NASA Conference Publication 2350

Proceedings of the Workshop on Improvements to Photometry

November 1984

NASA
National Aeronautics and Space Administration

Proceedings of a workshop held at
San Diego State University
San Diego, California
June 18-19, 1984

NASA Conference Publication 2350

Proceedings of the Workshop on Improvements to Photometry

Edited by W. J. Borucki and A. Young

Proceedings of a workshop held at
San Diego State University
San Diego, California
June 18-19, 1984

NASA
National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

TABLE OF CONTENTS

between pairs of stars compared with each other on different nights, using very similar equipment. It is not clear how much of this unexplained variation is stellar (Lockwood), how much is atmospheric (Fossat), and how much is instrumental (Rosen).

In this discussion, terms such as "absolute" and "relative" photometry are insufficiently precise. We must be careful to specify quantitatively the angular, temporal, and wavelength dependence of the stellar or instrumental variations under study. Different astrophysical problems (stellar seismology; photometric searches for planets; stellar classification and galactic-structure studies) involve variations on very different scales in these quantities. Thus, whereas the "d.c." components of different stars cannot yet be compared at the millimagnitude level, short-period variations much smaller than a millimagnitude in amplitude have been measured in several stars.

The areas that need work are the following:

Stellar variations and constancy- Do stars exist that are constant to a millimagnitude on time scales of weeks, months, and years? Which stars are most nearly constant? We do not even know the true level of variability of the 500 brightest stars.

Atmospheric limitations- At a good site, the atmospheric extinction is constant enough that extra-atmospheric extrapolations can be reliably repeated to 0.003 mag from measurements of a single star--the Sun (Angione). If both rising and setting stars are observed simultaneously, slow changes in extinction can be measured and removed (Rufener). However, there is evidence of serious transparency noise on time scales longer than 10 min (Fossat).

We must know more about the angular and temporal scales of transparency variations, which are primarily due to aerosols. Are these the cause of the lower night-to-night repeatability of star-to-star comparisons, or are instrumental variations at fault? How much cancellation of transparency variations can be obtained by simultaneous observations of two (or more) stars, as a function of their angular separation? Can this be done with a multi-channel instrument, or are optical and detector instabilities so severe that rapid chopping must be used with a single-channel instrument?

Instrumental problems- As was just indicated, we need new techniques, both to obtain and use data at high temporal and angular resolution, and to model these data correctly to avoid introducing errors in data reduction (see below). Optical fibers offer an alternative to Fabry lenses (Fantone). Thermal and other problems in filters will probably be limiting at approximately an order of magnitude beyond current precision; some spectrometric approach may be required (Young). Instrumental temperatures could probably be measured more accurately than they can be controlled, and then used in the data reductions as independent variables, as was suggested by Rosenblatt (Icarus, 14, 71, 1971).

More data are needed on the angular, temporal, and wavelength variations of the telescopic point-spread function (Ken Knight). Simultaneous imaging of several stars allows more efficient use of the data by PSF-matching in data reduction (Djorgovski).

Detector stability is a controversial problem. There is agreement that actual measurements at the telescope show 1/f-like noise on time scales longer than 10 min that is neither photon noise or scintillation noise (which are both white on the time scale of interest). Fossat thinks it is mainly atmospheric; Rosen thinks it is mainly due to photomultiplier gain changes.

Solid-state photodiodes need to be studied for their long-term stability, as potential detectors for multi-channel instruments. Although the silicon diodes have been studied the most, Schaefer and Geist suggested that GaAs or GaP diodes may be more suitable for astronomical work. CCD cameras, which have mainly been used to detect very faint objects, need further investigation for their potential use in high-precision work (Djorgovski). The Cochrans emphasized the demonstrated virtues of Reticons at high s/n, and pointed out that their linearity and repeatability minimized data-handling problems. Other detector arrays, such as CIDs and MAMAs also need investigation.

General problems of photometric systems. Current algorithms used to model atmospheric extinction for data reduction are marginal, if not inadequate, even at present precision (Young). The atmosphere, telescope, instrumental optics, and detector form a system that must be modeled accurately in the data-reduction software. Part of this modeling requires the derivatives of the instrumentally smoothed stellar spectra to be close to each pass band of the photometric system. These can be obtained if the spectrum is adequately sampled. Since all existing photometric systems are grossly undersampled, we should observe the requirements of the sampling theorem, so that the necessary information can be obtained in the modeling for data reduction, thus providing the means for an extension of the limits of photometric precision (Young).

The more that can be learned about the appropriate data, the better the result will be. When we make a photometric measurement, we obtain a value for the flux of radiation entering the detector. The observed flux is an integral over time, wavelength, and area. This integration helps to improve the signal-to-noise ratio when the flux is small; however, the integration may have a deleterious effect on the photometry. To get around this effect, the Geneva photometry (see the paper by Rufener in these proceedings) cleverly uses the statistics of short integrations to understand the significance of the final result. This is essentially making an observation that is differential in time. Such a procedure is most effective on brighter objects since at the limit of detection one cannot afford to subdivide a small number of photon counts--this limit, however, is very faint even for a small telescope.

A second kind of differentiation involves the image. Several speakers at this workshop argued for photometry by imaging detectors. This has many advantages: the entire distorted and displaced image of the stellar object can be monitored as a guide to seeing conditions; the background sky brightness can be simultaneously observed to guard against interference of many kinds. The infrared astronomers have led the way in rudimentary methods for simultaneous viewing of object and background by "chopping" or "nodding." Within the last few years many new kinds of electronic image-sensitive detectors that provide full information in the visible range have made their appearance; Djorgovski (at this workshop) has given a lot of information about CCDs, which are now in common use at most major observatories. Walker comments that such detectors are inherently as stable as phototubes for normal stellar photometry; it is quite clear that the differential advantages of using the image to

improve the photometry have not yet been achieved in the same sense that the time series has been exploited by the Geneva photometrists.

Finally, a third kind of differentiation involves wavelengths. Young's comments on filters are interesting in this respect. The presence of colors (including abrupt spectral features) in the source, in addition to color dependences in the detectors, makes it difficult to intercompare different objects or even to do time-series photometry if there are variations. Simultaneous narrow-band spectral measurements would provide additional information--again on brighter objects--which could improve the photometric precision.

Recommendations

The workshop participants divided the work that must be done into three categories: experiments that can easily be done in the near future without new funding; experiments that require somewhat more time and effort; and major investigations that require substantial funding.

Things that can be done immediately-

1. Measure the point-spread functions of several telescopes, and look for variations with wavelength, air-mass, angular separation, and time.

2. Define the limitations of various detectors, especially their temporal variations. How closely will two detectors reproduce the variations in the same light source? This especially needs to be done with solid-state photodiodes.

3. A literature search needs to be done to find out what is known about intensified arrays, CIDs, MAMAs, etc., as high-precision detectors.

4. Wes Lockwood will complete his analysis of the Lowell data to try to locate the source of the 0.003-mag limitation.

Time-consuming elements that need to be considered- Determine the atmospheric limitations more precisely, as described above. (Someone should also determine the conditions for photometry during the South Polar night, but this is more ambitious.) Two telescopes at the same observatory should observe the same stars simultaneously, to eliminate stellar variations as a noise source.

We should determine the precision of existing photometers in a laboratory that is free from stellar and atmospheric effects. This should determine how much of the various and how instrumental they could be on time scales longer than 10 min; in combination with the previous measurements, it should also determine the relative importance of instrumental, atmospheric, and stellar variations.

The ultimate photometric precision of existing CCD cameras needs to be determined.

Items that need substantial funding- The current systematic errors must be eliminated from the models used to represent atmospheric extinction. More accurate models are needed for instrumental effects (e.g., can the effects of temperature and humidity changes on instrumental transmission be used to improve data reduction?), and better techniques (selection and scheduling of extinction stars; multichannel

observations of sky, and more than one star at a time) need to be fully developed to cancel atmospheric and instrumental variations. Finally, the improved models need to be translated into algorithms and then into actual computer programs for data reduction. This is a task which involves significant effort.

Bands must be chosen to satisfy the sampling theorem, to provide information needed to model the instrument and atmosphere accurately. This requires attention to band profiles as well as effective wavelengths; new instrumental methods that avoid filters may be necessary. These efforts must be translated into hardware if the improved modeling methods are to be used, because aliasing effects are already substantial at present precision.

An instrumental system that exploits the advantages of solid-state photodiodes should be developed. Laboratory tests of the limits to photometric precision of such diodes are needed, including studies of long-term stability. If such a system is correctly sampled, and properly modeled in the data-reduction software, it can be used to determine the true variability of the 500 brightest stars. This would provide valuable statistics on stellar microvariations.

These experiments will allow a fuller investigation of the temporal, angular, spatial, and spectral dependences of the instrumental, atmospheric, and stellar variations. Those results are needed to design photometric systems intended to reach the ultimate limits of photometry, and to decide whether photometric methods can be used for such demanding tasks as planetary-system detection.

Communication

Many of us were surprised to find so many other people with similar or related interests. Instead of being a handful of isolated workers, perhaps regarded as a bit peculiar by their colleagues, we discovered that we really are a community of high-precision photometrists.

After some progress has been made, another meeting on high-precision photometry to discuss significant results and to decide what appropriate steps are to be taken would be valuable. This community enthusiastically anticipates such a meeting to take place within the next year or two.

I. ASTRONOMICAL APPLICATIONS OF HIGH-PRECISION PHOTOMETRY

N85
17893

UNCLAS

HISTORY OF PHOTOMETRIC PRECISION

Andrew T. Young

San Diego State University

INTRODUCTION

The history of astronomical photometry was beautifully reviewed by Weaver (1946), in a series of articles that should be read by all serious photometrists. This history is a long one, almost as long as that of astronomy itself; for even the oldest surviving star catalogs contain estimates of the apparent brightnesses of the stars, as an aid to their identification.

In the Almagest, we not only find the naked-eye stars divided into six classes of "magnitude" (a translation of the term used by Ptolemy), but also see these divided into "greater" and "lesser" subdivisions, which were subsequently interpreted as thirds of a magnitude. Thus, the beginning of a quantitative numerical scale was already established nearly 2000 years ago by Hipparchus and Ptolemy. As Weaver points out, modern magnitude scales are directly descended from those determined by the ancient Greeks.

FIXING THE SCALE

There is, however, an important distinction that should be made. The ancient visual magnitudes formed a perceptually uniform scale: that is, they represented equal steps in visual sensation. The modern magnitude scale (Pogson, 1857) is a logarithmic one. Although the divergence of the logarithm function at zero argument obviously means that the eye's response cannot be logarithmic (otherwise we would be overwhelmed by an infinite negative sensation in total darkness), this incorrect functional dependence was adopted in the mid-19th Century to fit the traditional visual scales to physical measurements of stellar intensities.

The functional relation between visually-estimated magnitudes and physical intensities was established only slowly, through several approximations. Newton assumed that 2nd magnitude stars were twice as far away as 1st, that 3rd magnitude stars were 3 times as far, and so on. This would make the brightnesses inversely proportional to the squares of the magnitudes, or the magnitudes the inverse square roots of the brightnesses, which is not too far from the truth (cf. Fig. 1). Yet it was far enough to deter Newton from publishing his repeated attempts to use this principle. A century later, William Herschel went through exactly the same arguments in public (Hoskin, 1977). Sir John Herschel found a slightly modified form of this law to be true from his photometric measurements made at the Cape of Good Hope; but by the time his results were published, Steinheil (1837) had introduced the logarithmic relation, which was adopted by nearly all later workers.

An historian of science would say that the universal use of logarithms for astronomical calculations had predisposed the astronomers to adopt a logarithmic scale. Indeed, Steinheil (1837) assumed such a scale for convenience,

and F.G.W. Struve seems to have made the equivalent assumption that one magnitude step corresponds to a fixed ratio of brightnesses. In any case, the actual nature of such scales is more correctly described by a power law (Stevens, 1961, 1975). Thus, Weaver's assumption that logarithmic magnitudes are "right" and the older visual scales were "wrong" is mistaken.

In fact, if one plots the logarithms of Ptolemy's magnitudes against the logs of the corresponding stellar brightnesses (based on modern visual magnitudes), one finds (see Fig. 1) that the magnitudes in the Almagest form a remarkably linear (power-law) scale. The systematic error that Weaver describes as the "scale error" in the old visual scales is in fact due to the adoption of an inappropriate functional form in the 19th Century; the error is due to Steinheil and Pogson, not to Hipparchus and Ptolemy. One should bear this in mind when reading Weaver's discussion of systematic errors in these older scales.

STEPS IN PRECISION

Allowing for this defect, we can now examine Weaver's discussion of the precision (which he calls "Internal accuracy") of various observers' work. Figure 2 shows how this has improved with time. I have added a few recent data to supplement those in Weaver's tables.

The remarkable feature of this figure is the agonizing sloth with which photometry has improved over the years. I cannot think of any other physical measurement that has become only two orders of magnitude more precise in two thousand years, and has improved by little more than a factor of ten in the past century. The very first photoelectric photometry, which was done before World War I, was of about the same quality that is routinely achieved today. What was once an arcane art, practiced only a handful of experts, has become a standard technique; but though it is now done more easily, it is not done much better. Why?

Perhaps a survey of historical progress will suggest the kinds of difficulties that beset the photometrist, and help suggest areas to examine for possible future progress. The first big advance, of roughly a factor of three in precision, was simply due to being careful, and explicitly attending to photometry as a study worthy in itself, instead of a mere adjunct to stellar identification.

The next big step was the introduction of optical assistance in the visual comparison of one star with another. The visual photometers dimmed the light of a bright star until it visually matched that of a fainter one. The eye was thus used only to detect equality or inequality, not to estimate differences. This produced only about a factor of three improvement, reducing typical errors from about 0.2 mag to 0.06 mag.

The next gain came with the introduction of photography; but I think it is a mistake to view this (as is usually done) as an improvement in detectors -- though this was the way it was viewed at the time. For example, Bond (1859) expected that photography "would have an unquestionable advantage" over the visual techniques then used: "it would meet perfectly the greatest of the many

difficulties of the problem -- the comparison of stars exhibiting diversity of color."

Subsequently, however, the photographic plate proved to have all the faults of the eye: variable sensitivity, nonlinearity, wavelength dependence of all its properties, and a complicated time-dependence (reciprocity failure) that is essential to any materials that are not quickly fogged black by integrating the thermal background. Indeed, similar faults (though at a subtler level) have now been discovered in CCD's, for which we had recently heard optimistic claims sounding very like those made a century ago for photography.

One must realize that the actual "detector" in much of the photographic work was still the human eye. Although Pickering suggested in 1910 that the thermopile or other "physical" detectors might be used to measure plates, much of the advance to be gained by using photography had already been gained by then. Rather, the advantage of photography was largely its ability to integrate and store the light signal. The long photographic exposures completely average out the scintillation noise, which even Steinheil had remarked as a problem in visual work; and the stored image allowed photographs to be "measured" (still mainly by eye!) at leisure, by comfortable and rested observers, instead of in real time by a cold, tired astronomer at the telescope. This also produced about a factor of two.

Also, the replacement of the eye with photography was a distinctly retrogressive step in one respect: the detective quantum efficiency of the early plates was one or two orders of magnitude worse than that of the eye. Even the best plates today remain slightly inferior to the eye in this respect. Yet no one today would think of the eye as a "better" photometer -- though it is, in this particular sense. The lesson here is that detectivity isn't everything.

Though photography has continued to improve, and in some senses (Latham, 1978; Davis et al., 1980) remains competitive with photoelectric detection, the photoemissive detectors certainly alleviated a major photographic problem: nonlinearity. Its practical consequence is that one must take enough calibration data to define the nonlinearity to the desired precision. But if one wants high precision, one must devote an enormous effort to determining the calibration curve -- a problem that is just as severe with "modern" detectors like TV systems. Usually, either (a) the user isn't willing to take the necessary care in calibrating these nonlinear detectors, and then blames the bad results on the nonlinearity; or (b) the calibration time eats into observing time so that little advantage is gained, because of inefficient use of the telescope. (This is becoming a problem with CCD's, as their quirks become better known.)

Perhaps a more fundamental advantage of photoelectric systems was their improved accuracy of integration, without the complicating effects of a characteristic time scale. (In photography, this is the exposure time at which the emulsion is most sensitive, and obeys reciprocity.) This allowed time-scale photometry to become a respectable technique at last, and helped extend the limited dynamic range of the early electrometers.

Finally, photocells of all types removed a human element that remained in photography: the nonuniformities due to uneven agitation in photographic processing. (Photomultipliers have their own counterpart to this problem of reproducibility: gain drift, or "fatigue".) Careful laboratory studies (Latham, 1966, 1968) have shown that these large-scale nonuniformities can be reduced to one or two per cent, which is about the level of error reported in the best photographic work. Although an additional factor of 10 can be reached by Latham's "grid photography" technique, so much labor is required that it seems not to have been tried by anyone else.

Together, these advantages of photoelectric detectors secured a marked improvement, of about a factor of 5, in photometric precision.

Note that the higher quantum efficiency of photoemissive cells contributed nothing to this: it only allowed fainter stars to be measured in a reasonable time. In fact, signal/noise ratio has played only a minor part in this story. The central theme has been the detection and removal of various systematic effects; particularly, those connected with the accuracy of averaging. Better S/N has helped reveal smaller effects, and thus only indirectly led to their correction.

DISCUSSION

Which of the historical problems are still with us? We still have problems with stars of different spectral energy distributions, because of the difficulty of maintaining and reproducing our instrumental spectral response. We still have a human element in the centering of stars; because our optical systems do not have uniform responsivity with respect to solid angle, this may remain a problem.

Certainly, one reason for the difficulty of measuring light is its vectorial nature. The orientations of both the Poynting vector and the electric vector with respect to all optical surfaces, including that of the detector, affect the instrumental sensitivity, and hence the measured value. Yet, these effects are today only of minor importance, and cannot have been significant in the past.

Some are tempted to blame our detectors. Yet,photomultipliers have been used in the laboratory to detect differences smaller than one part in 10,000; and similar precision has been obtained in stellar polarimetry. This is far beyond the few parts in 1000 that seems to be the limit of photometric precision today (cf. Fig. 2). Certainly, it appears that present detectors should be able to do substantially better than they have done. The history ofphotometry is replete with glowing claims for detectors that have not lived up to the promises of their advocates. I think that optical problems connected with the spectral and angular response of the instrument are the probable limitations at present.

Nevertheless, it is by no means clear that the present limitations of photometry are either understood or accounted for. That is why we are here: to examine these questions, and then to decide what to try next to improve photometric precision.

REFERENCES

Bond, G. P. (1859). A. N. 49, 81.

Davis, M., Feigelson, E., and Latham, D.W. (1980). A. J. 85, 131.

Hoskin, M. A. (1977). J. Hist. Astron. 8, 77.

Latham, D. W. (1966). Astron. J. 71, 168.

Latham, D.W. (1968). Astron. J. 73, 515.

Latham, D.W. (1978). "Signal-to-noise", in Modern Techniques in Astronomical Photography, eds. R. M. West and J. L. Heudier (ESOm Geneva), 141-149.

Pogson, N. (1857). M. N. 17, 12.

Steinheil, K. A. (1837). Abh. math.-phys. Cl. Bayer. Akad. Wiss. 2, 3.

Stevens, S. S. (1961). Science 133, 80.

Stevens, S. S. (1975). Psychophysics (John Wiley & Sons, New York).

Weaver, H. F. (1946). Pop. Astr. 54, 211, 287, 339, 451, and 504.

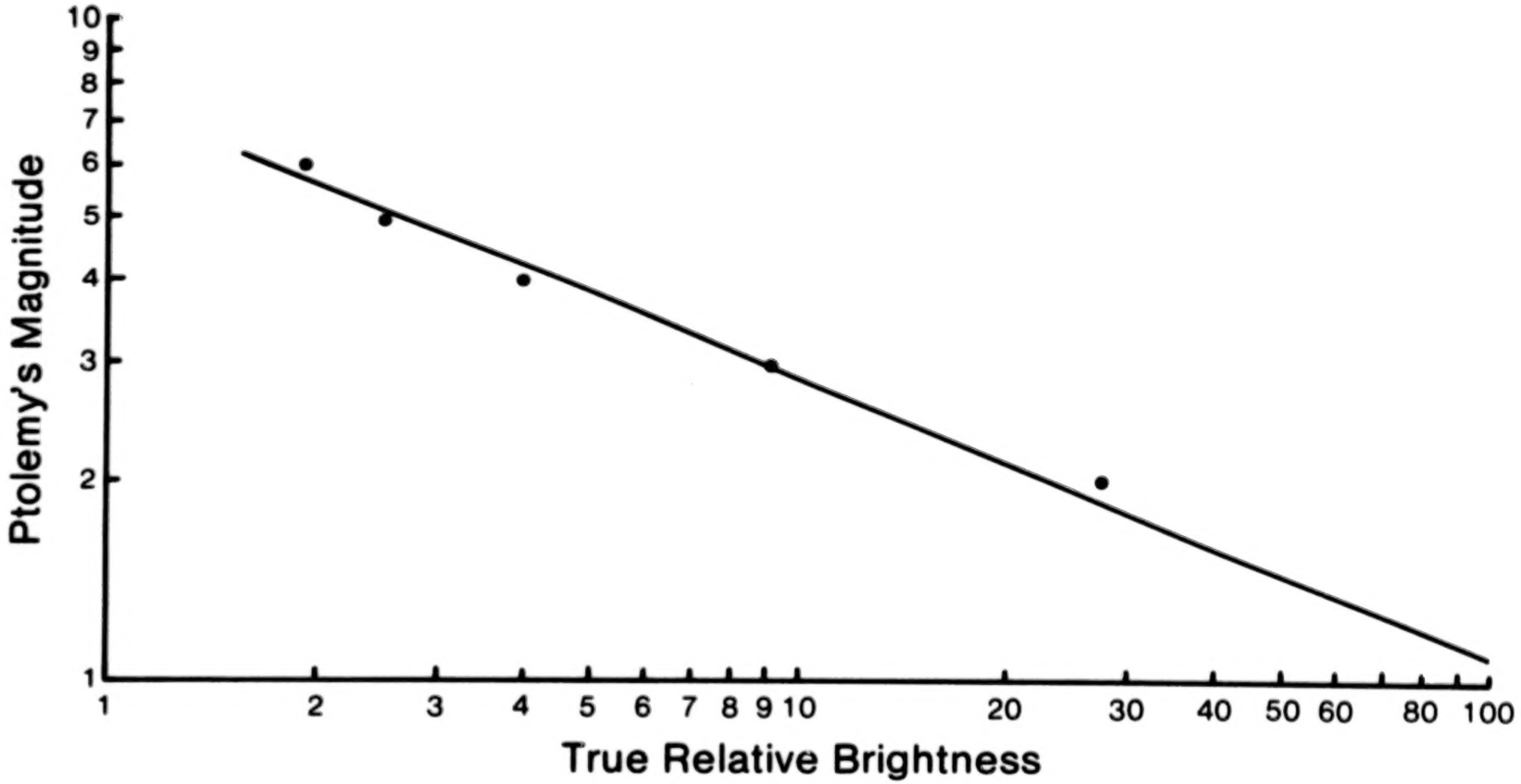


Figure 1. Logarithms of Almagest magnitudes plotted against the logarithms of the average brightnesses of the corresponding stars (taken from Weaver's Table I.)

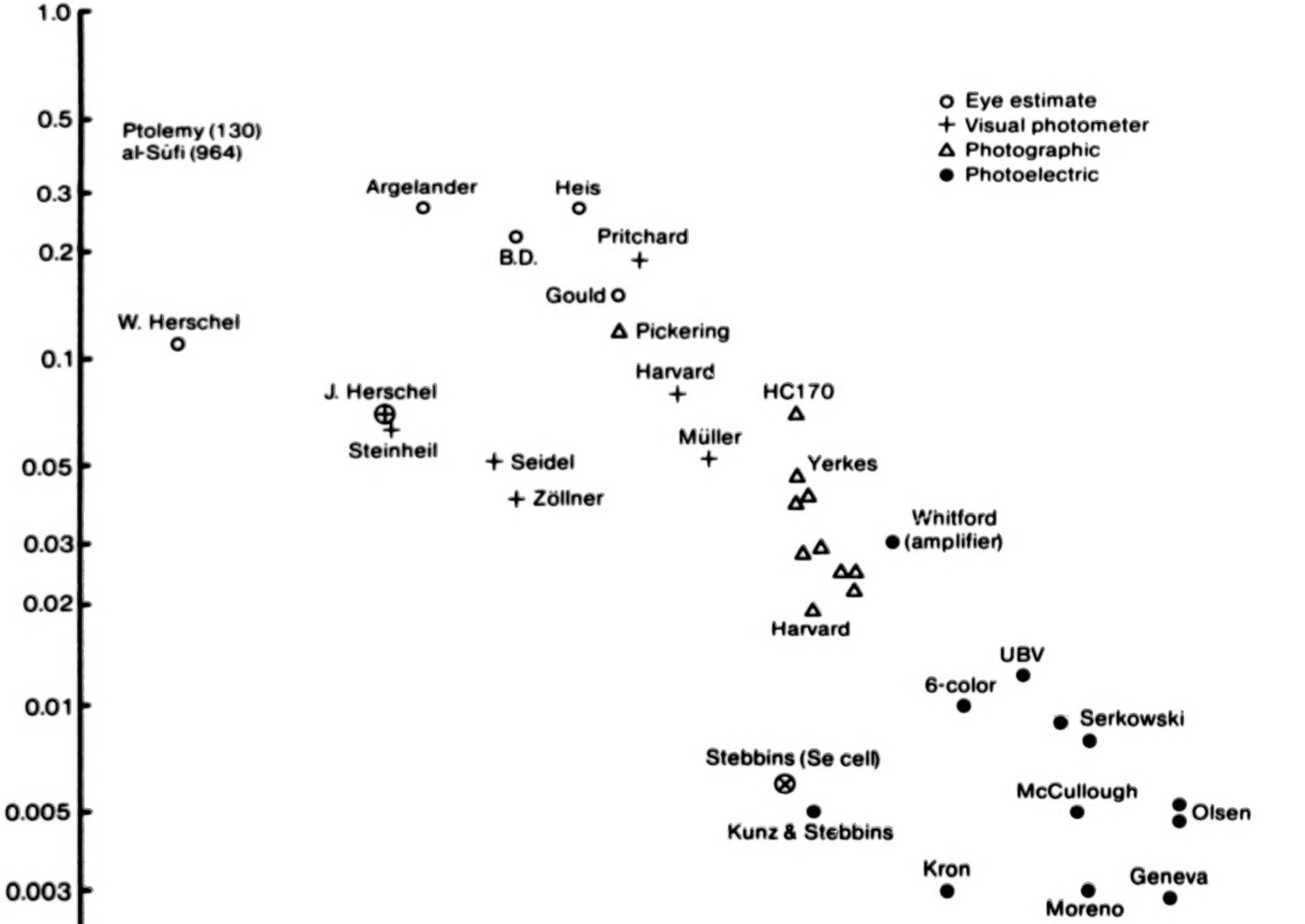


Figure 2. Internal errors of various careful photometric measurements, as a function of time. The earlier data are taken from Weaver (1946); recent ones have been added.

D2

N85-17894

PHOTOMETRIC PRECISION NEEDED FOR PLANETARY DETECTION

W. J. Borucki

Ames Research Center

SUMMARY

The photometric method of searching for planets around stars depends on observing the decrease in light flux that is produced by the transit of a planet across the stellar disk. The magnitude of this reduction is proportional to the ratio of the planet's area to that of the star. For the solar system, the decrease in light amounts to 1.0% for giant planets such as Jupiter and Saturn, 0.1% for planets such as Uranus and Neptune, and 0.01% for Earth-sized planets. The photometric method works only for planets whose orbital plane includes the Earth. Consequently many stars must be monitored on a nearly continuous basis.

Aside from technical limitations imposed by the photometer and the fundamental limit of photon counting statistics, the intrinsic photometric variability of stars imposes a practical limit on the detectability of small planets. Based on recent precise solar observations made from the Solar Maximum Mission (SMM) satellite, it appears that the detection of terrestrial-sized planets will be marginal during periods of high stellar activity, but that the detection of such planets during quiescent periods appears possible.

To overcome the effects of scintillation and differential extinction, it will be necessary to use large aperture telescopes and narrow-pass spectral filters, or to operate the photometric system on a space platform. A state-of-the-art photometer is being developed as a test-bed to test components and concepts. The goal is the development of a photometer that can routinely measure the relative brightness of stars to a precision of 1 part in 10^5.

INTRODUCTION

Detection of extrasolar planetary systems is important to the understanding of star formation, the formation of planetary systems, and to the estimation of how common life is throughout the universe. Black (1980) presents a summary of the various techniques that have been considered for detecting other planetary systems. Efforts are under way to implement the radial velocity and astrometric techniques. As presently envisioned, these techniques will be able to detect major planets that are at least as massive as Uranus and Neptune, but will not be able to detect planets as small as the Earth. Rosenblatt (1971) and Borucki and Summers (1984) discuss photometric methods that have the potential of detecting Earth-sized planets if a photometer with the required precision can be developed.

The simplest photometric method of searching for planets around stars depends on observing the decrease in the starlight that is produced by the transit of a planet across the stellar disk. The magnitude of this reduction is proportional to the ratio of the planet's area to that of the star. For the solar system, the decrease in light amounts to 1.0% for giant planets such as Jupiter and Saturn, 0.1% for

planets such as Uranus and Neptune, and 0.01% for terrestrial-sized planets. The photometric method works only for planets whose orbital plane includes the Earth. Because the probability of observing a planetary transit during a single observation of a randomly chosen star is small, the search program must be designed to continuously monitor many stars so that a discovery rate of 1 planet per year can be attained. Two different approaches appear promising; monitor approximately 100 stars with a precision of 1 in 10^5 to detect small, inner planets with short orbital periods, or monitor approximately 10^4 stars with a precision of 1 in 10^3 to detect giant outer planets with long orbital periods. In this paper only the first option will be considered. See Borucki and Summers (1984) for a discussion of the second option.

Aside from the technical limitations that are imposed by the photometer and the fundamental limit of photon counting statistics, the intrinsic photometric variability of stars imposes a practical limit on the detectability of small planets. No star can be regarded as completely nonvariable. Small variations of stellar flux are caused by convection and star spots that are modulated by the rotation of the star. Based on recent precise solar observations made from the Solar Maximum Mission (SMM) satellite, it appears that the detection of Earth-sized planets will be marginal during periods of high stellar activity, but that the detection of such planets during quiescent periods appears possible. The detectibility of planetary transits in the presence of solar-type noise is treated in Borucki et al. (Detectability of Extrasolar Planetary Transits, in press, 1985).

This paper presents an overview of the photometric method of detecting other solar systems, the precision required to detect planets of various sizes, and it discusses the limitations to the available precision caused by stellar variability, atmospheric variations, and instrument problems. A sketch is presented of a prototype photometer to test components and concepts.

PROBABILITY OF OBSERVING A TRANSIT

To estimate the expected planetary detection rate, it is necessary to estimate the probability of observing a star that is being transited by a planet. Let p be the probability that the orbital plane is properly oriented so that transits can be observed, let d be the stellar diameter, r the orbital radius, G the universal gravitational constant, M be the mass of the star, τ the orbital period of the Earth, and let n be the number of transits per star that are expected to be observed in one terrestrial year. In Borucki and Summers (1984) it is shown that

$$p = d/2r \quad \text{and} \quad n = \frac{d\sqrt{GM}}{4\pi r^{5/2}} \tau \qquad (1)$$

Equation (1) shows that n is largest for close-in planets of large and massive stars. For equation (1) to be valid, observations must be carried out 24 hr per day, i.e., 3 or 4 telescopes equally spaced in longitude must be used or the observations must be carried out from space. If observations are carried out for an average of 6 hr per day, then p and n should be reduced slightly for inner planets that have a transit-time of less than 18 hr (see Borucki and Summers, 1984). Let it be assumed that a photometric precision of 10^{-5} can be obtained and that this is sufficient to detect planets the size of Venus and larger. Table 1 gives n and its summation N,

for Venus, Earth, Jupiter, Saturn, Uranus, and Neptune using the assumption that on the average, every star has a solar system like our own.

It is clear from table 1 that if 65 (i.e., the reciprocal of 1.53×10^{-3}) suitable stars can be monitored continuously, then the detection of 1 transit per year can be expected. Because the highest probability of an observation is associated with planets with short orbital periods, a confirmation of a planetary detection could be made within a year by detecting a signal of the same amplitude and duration one orbital period later. The orbital period can be predicted from the duration of the transit and from an estimate of the stellar diameter. The confirmation of transit observation can be strengthened by using two-color photometry. See the discussion in Borucki and Summers (1984). By observing stars in open clusters, it should be possible to monitor hundreds of stars and thus detect 5 or 6 transits per year.

Although space-based radiometric measurements of the sun have attained a relative precision for hour-long observations of 10^{-5} (Wilson et al., 1981), current stellar photometers that must observe much dimmer objects through the terrestrial atmosphere obtain a night-to-night precision of only 3×10^{-3}. Three sources of error can be identified; variability of the star itself, variations in the flux due to atmospheric variability, and instrument problems. Each of these noise sources will be discussed in the following sections.

ESTIMATE OF STELLAR VARIABILITY

To simulate what a signal from a planetary transit would look like in the presence of a stellar variability, data (Wilson et al., 1981) from the 1980 observations of total solar irradiance by the radiometer aboard the SMM satellite were combined with calculated flux reductions caused by a planetary transit of the Sun by Jupiter and by the Earth. Figure 1 shows what a distant observer in Jupiter's orbital plane would observe during such a transit. The Jupiter signal is clearly detectable. A simulation of the signal that is expected from the transit of an Earth-sized planet across the Sun is presented in figure 2. The points on the solid curve represent 1-hr samples of the solar flux measured by the SMM radiometer. The dashed curve connects points that represent the flux that is expected if an Earth transit had occurred during the measurements. The data give a S/N of approximately 4, but it is clear that the variability of the Sun makes the detection of terrestrial-sized planets marginal. The SMM observations shown in figures 1 and 2 were made at the peak of the solar activity cycle. The solar variability for other portions of the solar activity cycle is expected to be much less than that shown in figure 2. Therefore, detection of terrestrial-sized planets appears possible when the stellar activity levels are below the peak activity level of the Sun. The detection of larger planets and low-mass companion stars should present no fundamental problem even in the presence of substantial stellar activity levels. However, it is clear that stellar variability is a fundamental limit to the ability of the photometric method to detect small planets.

A variation of the photometric method, two-color photometry, uses simultaneous observations through short-wavelength and long-wavelength filters to reduce the sensitivity of the observations to starspots and to fluctuations in atmospheric extinction (Borucki and Summers, 1984). The increased ability of the two-color method to discriminate against noise sources is obtained only if the measurement precision of the color ratio is approximately 5 times greater than that required to measure the

fluxes directly. In summary, it is clear that the precision of photometric measurements must be at least 1 part in 10^5 but need not exceed 2 parts in 10^6.

Because transit times for Earth-sized planets range from 11 to 16 hr for the orbital periods that are appropriate to Venus, Earth, and Mars, the duration over which the precision must be maintained is approximately 32 hr, i.e., twice that of the longest expected duration. The longest sample-time that can be considered is 11 hr. If there is a desire to search for planets with orbital periods as short as that of Mercury, then the maximum sample time is only 8 hr rather than 11 hr. In the next two sections of this paper, the consequences of atmospheric and instrument limitations will be considered in light of the required precision to detect Earth-sized planets.

ATMOSPHERIC LIMITATIONS

Because the terrestrial atmosphere imposes severe limitations on the available photometric precision, a ground-based measurement program will require a much larger aperture telescope than a spaceborne program would require.

Two major sources of signal degradation are fluctuations in terrestrial atmospheric extinction and scintillation. (Seeing does not appreciably affect differential photometric measurements if large focal-plane apertures and careful techniques are used.) To reduce the effect of extinction fluctuations, Radick et al. (1983), used an ensemble of reference stars instead of using a single reference star. However, they found that such a procedure was not capable of producing a night-to-night precision better than 0.004 because of fluctuations in atmospheric extinction. Grauer and Bond (1981) and DeBiase et al. (1978) have shown that the use of a two-channel photometer which simultaneously views two stars can dramatically improve the precision of the measurements. Figure 3 was adapted from a paper by Geyer and Hoffmann (1975), and shows the cancellation of extinction fluctuations that they observed when using a two-channel photometer. They state that on nights when cirrus were present, the two-channel photometer reduced fluctuations of 50% to less than 1%.

After first-order extinction has been removed by simultaneously measuring the ratio of the fluxes of two stars, there will still be a residual error because of differential extinction unless the two stars have identical spectral curves in the spectral pass band of the filter. For quantitative discussions of this effect, see King (1952) and Young (1974). Calculations were made to estimate the change in the apparent flux ratio of two main-sequence dwarfs that differ by one spectral class caused by the air mass changing from 1 to 2. The stellar spectra were approximated by grey bodies with color temperatures given in Allen (1963). The filter response r, was approximated by using a sine-squared response. The extinction versus wavelength curve presented by Hardie (1962), for Mount Wilson and McDonald Observatories was used to calculate the extinction α versus wavelength λ for varying air masses M. The ratio R of the two fluxes was calculated from equation (2).

$$R = \int_{\lambda_1}^{\lambda_2} r\, e^{-\alpha M}\, B(\lambda,T_1)\,d\lambda \Big/ \int_{\lambda_1}^{\lambda_2} r\, e^{-\alpha M}\, B(\lambda,T_2)\,d\lambda \qquad (2)$$

where $B(\lambda,T)$ is the Planck function and T is the color temperature of the star. Graphs of the fractional change in the flux ratios versus filter width are presented

in figure 4 for stars of various spectral class and for two wavelengths. The 500- and 700-nm wavelengths refer to the center bandpass of the filters. From figure 4, it can be seen that the fractional change of R is proportional to the square of the filter width and is quite sensitive to the center wavelength, but it is nearly independent of the spectral class with the exception of M-class stars. It is clear from figure 4 that R must be corrected for air mass to attain a precision of 10^5 and to use filters with center wavelengths as short as 700 nm and widths in excess of 15 nm. The use of narrow band filters requires long integration times. The properties of the filters must be very stable because the spectral position and width of the bandpass are sensitive to temperature. Although interference filters are much less sensitive to temperature than are glass filters, they degrade with time because of water vapor absorption. Both types of filters have spatial nonuniformities and therefore require special care to ensure that the position and direction of the light beam that passes through the filters do not change. (See the discussion on filter properties by Andy Young in this volume.)

For bright stars, scintillation noise will dominate shot noise. Young (1974) showed that scintillation noise can be expected to vary with telescope aperture to the -2/3 power, integration time to the -1/2 power, the wind speed, and with the square of the air mass. In particular, for a sample time of 8 hr and an air mass of 2, a telescope with a 9-m aperture is required to keep the scintillation noise at, or below, 1 part in 10^5.

Because of the effects of extinction and scintillation caused by the Earth's atmosphere, only telescopes of very large aperture would have a chance of detecting Earth-sized planets. Observations from a space platform would eliminate the scintillation noise and would allow the elimination of the band-pass filters. Consequently the aperture of a spaceborne telescope could be substantially smaller than that of a ground-based telescope. A space-based telescope would also have the advantage of being able to make measurements 24 hr per day.

The impact of operation in the terrestrial atmosphere can best be seen by comparing the system required to obtain a precision of 10^5 while operating in the terrestrial atmosphere with that required for an operation in space. The solid line in figure 5 shows the dependence of the size of the telescope aperture on the star brightness for a ground-based system. Photon noise and detector noise overwhelm scintillation noise only for stars dimmer than 11 magnitude. To compute the solid curve shown in figure 5, it was assumed that the air mass was 2, that the spectral band-pass filter width was 20 nm, that the detector noise was 10^{-16} $W/Hz^{1/2}$ (Rieke et al., 1981), that the detector was a cooled silicon diode sensitive over the spectral range from 400 to 1100 nm, and that the combined quantum efficiency of the detector and the instrument throughput was 0.3. The dashed curve in figure 5 represents the telescope aperture of a spaceborne system with the same detector. Because of the absence of scintillation and the ability to make observations without a narrow band-pass filter, the spaceborne system can use a 1-m aperture telescope rather than needing a 9-m aperture to observe stars to $M_V = 11$.

INSTRUMENTATION FOR HIGH-PRECISION PHOTOMETRY

Before a spaceborne system is developed, it is important to determine whether there are ground-based approaches that can avoid the difficulties mentioned earlier and to develop a photometer that has the intrinsic precision required. A sketch of a test-bed photometer that could be used to test different types of detectors, preamps,

TABLE 1. NUMBER OF TRANSITS EXPECTED TO BE OBSERVED PER STAR PER TERRESTRIAL YEAR

Planet	
Venus	1.05×10^{-2}
Earth	4.65×10^{-3}
Jupiter	7.54×10^{-5}
Saturn	1.65×10^{-5}
Uranus	2.9×10^{-6}
Neptune	9.4×10^{-7}
Sum	1.53×10^{-2}

20-21

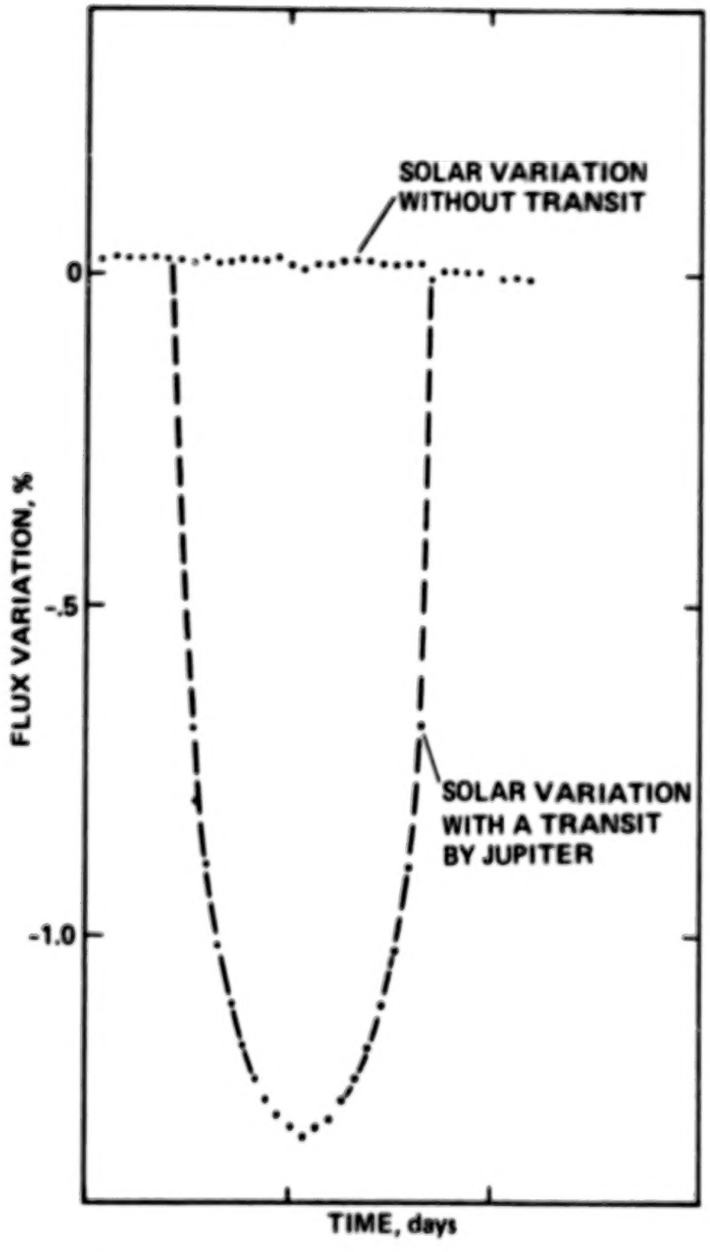


Figure 1.- Simulation of the flux variation of the Sun caused by a transit by Jupiter as seen by a distant observer.

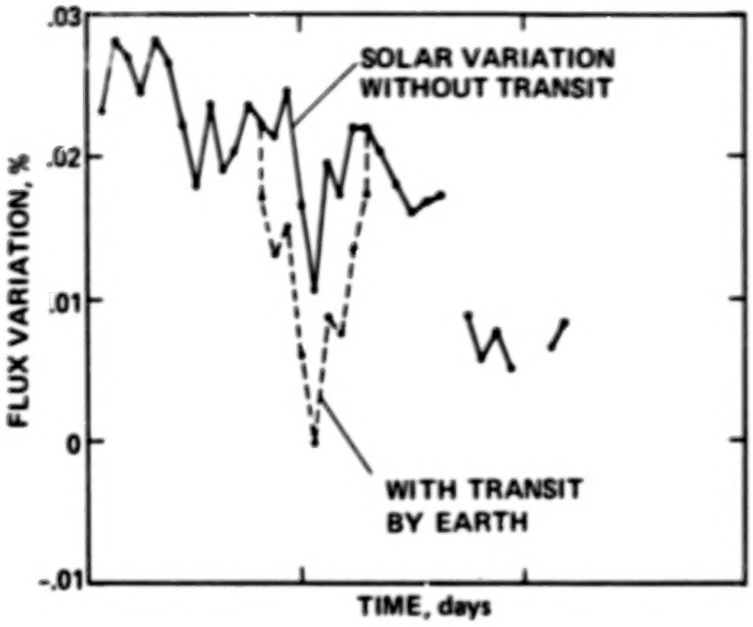


Figure 2.- Simulation of the flux variation of the Sun caused by a transit by the Earth as seen by a distant observer. The solid curves in both figures are measurements from the active cavity radiometer aboard the SMM satellite (Willson et al., 1981). Note that the scale on the ordinate is greatly expanded compared to that used in figure 1.

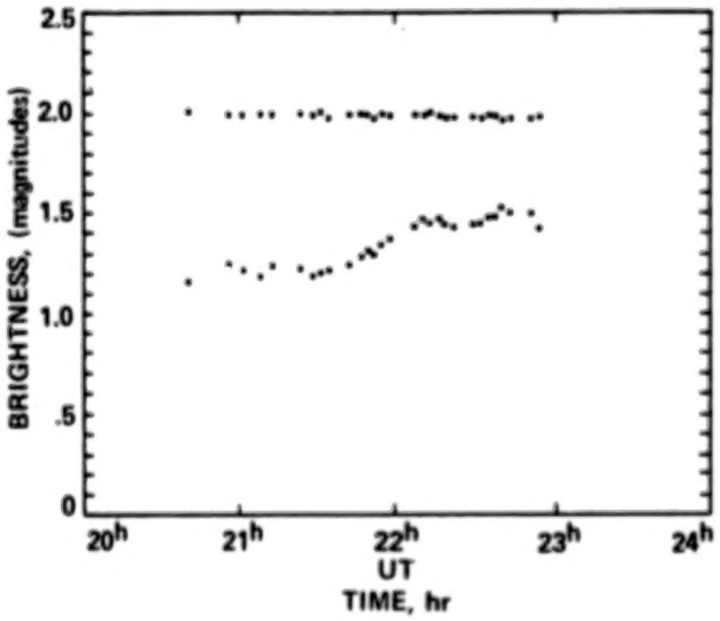


Figure 3.- The measured brightness and brightness differences (in magnitudes) for two stars as a function of time. The upper curve represents measurements by a two-channel photometer and the lower curve shows the measured brightness for one of the two stars. Adapted from Geyer and Hoffman (1975).

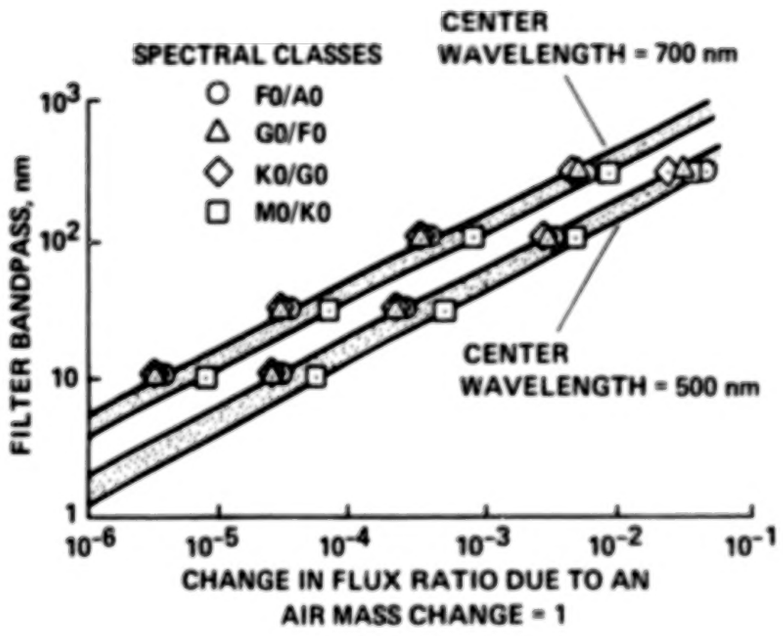


Figure 4.- Maximum acceptable width (nm) of the filter band pass for a specified value of the change in flux ratio for an air mass change equal to 1. Different symbols are used for the ratio for stars of different spectral classes. The upper stippled band is for filters with a peak transmission at 700 nm while the lower stippled band is for filters with a peak transmission at 500 nm.

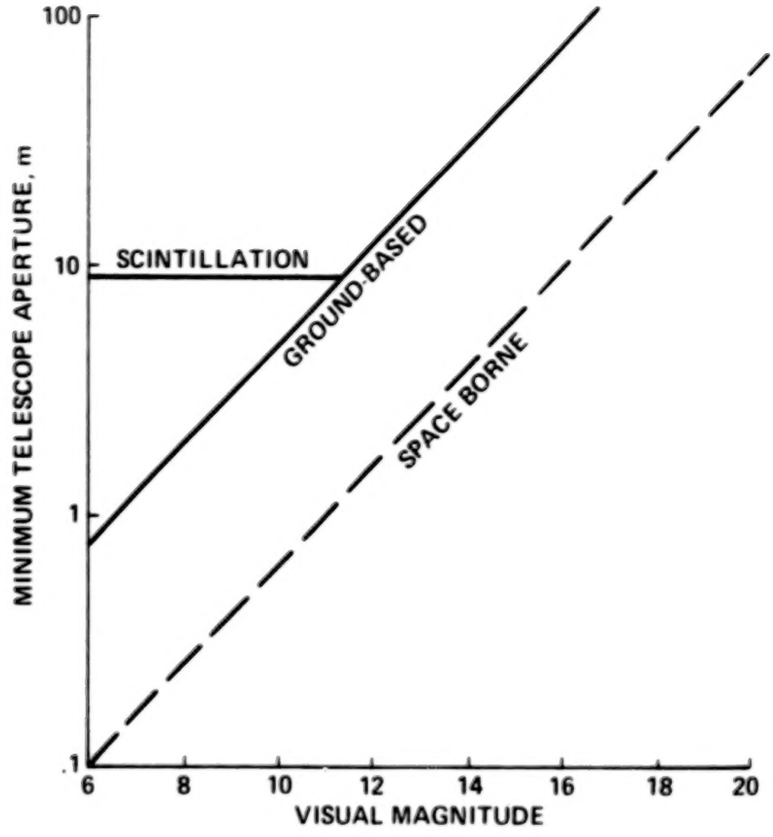


Figure 5.- Minimum telescope aperture required to produce a precision of 10^5 for an 8-hr sample. The horizontal portion of the solid curve represents the effect of scintillation. The rising portion of the solid curve represents the effect of using a detector with a NEP of 10^{-16} W/Hz$^{1/2}$, a filter with a bandwidth of 20 nm, and an air mass of 2. The dashed curve is for a spaceborne instrument using the same detector, but with no filter and a zero air mass.

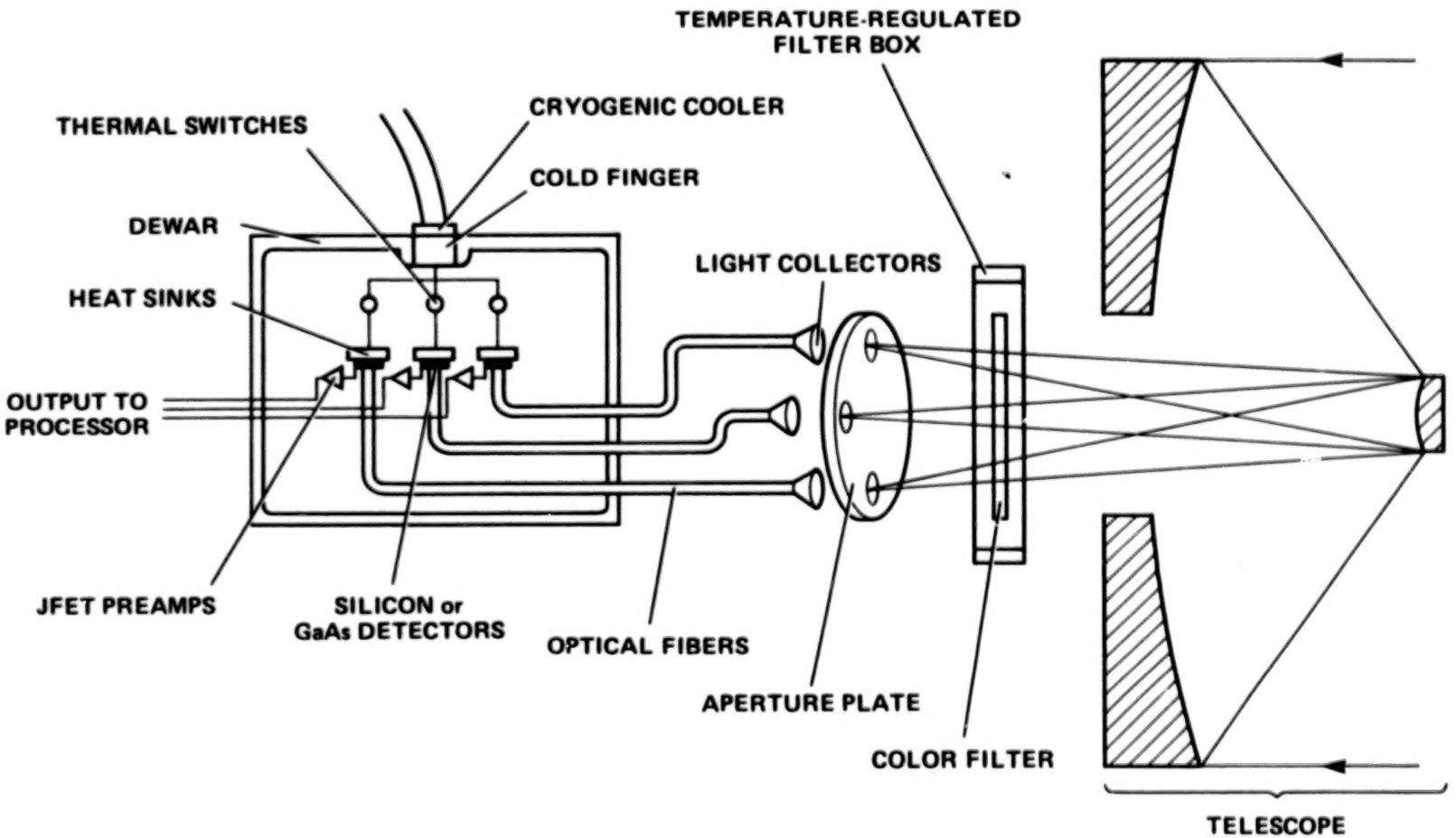


Figure 6.- Prototype photometer to test components and concepts. The goal is the development of a photometer that can achieve a relative precision of 1 part in 10^5.

FIVE-MINUTE P MODES DETECTED IN DOPPLER SHIFT MEASUREMENT ON ALPHA CENTAURI

Eric FOSSAT, Gerard GREC, and Bernard GELLY
Observatoire de Nice and
Departement d'Astrophysique de
l'Universite de Nice
Parc Valrose
F - 06034 NICE CEDEX

Last year, a short note (Fossat et al, 1983) described the first test of a new spectrophotometer specially designed for extending to a few bright stars the results already obtained in solar seismology. Since the late seventies, we know that the Sun is pulsating within a certain range of harmonics, the most famous having periods around five minutes. Very striking results in this field have been obtained by the observation of integrated sunlight. Indeed in this case, the angular filtering is so severe that only radial and weakly non radial (degree ≤ 3) eigenmodes can be observed. Their number is limited enough to make the identification possible despite the lack of any angular resolution in the observation. More than 80 of such eigenmodes, attributed to the pressure acting as a restoring force, have been thus identified in the five-minute range in the case of the Sun (Grec et al, 1983). Once identified in angular degree, radial order and temporal frequency, these eigenmodes make possible a real seismological investigation of the internal structure of the Sun.

Because such important results have been obtained in integrated sunlight, observing the Sun "as a star", it was tempting to try to achieve similar results on other stars. Quite evidently, the 10^{11} flux reduction factor for the brightest stars makes the task highly difficult, because the oscillation amplitudes to be detected are less than 1 m/s in Doppler shift measurements. It is for this special goal that we have designed a spectrophotometer, using again the principle of optical resonance spectroscopy. The conclusion of the first test of this new instrument was that if the observation can be photon noise limited (i.e. in total absence of any instrumental source of noise), the five-minute solar oscillation could still be detected by removing the Sun far enough for its magnitude to become zero or one.

Such a situation is very closely represented by the observation of Alpha Centauri A, because it is a G2 V star, very similar to the Sun, with a mass of 1.1 in solar unit. Six nights were granted to this program on the ESO 3.6m telescope, 22 - 28 May 1983. Two and half nights provided over 20 hours of data of photometric quality good enough for analysis. In fact, this data consists of two signals:

- The monochromatic intensity (about 0.08 Å bandwidth) in the red wings od the Na D1 and D2 lines.

- A reference channel, which contains the whole 20 Å passband of the interference prefilter.

The first step of the analysis consists in dividing the monochromatic intensity by the reference signal, in order to minimize the effect of atmospheric transparency fluctuations as well as the effects of scintillation. This has proved to be sufficient in the presence of clouds absorbing as much as 60% of the normal flux. In presence of thicker clouds, the diffusion of the moonlight makes the division too inaccurate.

A harmonic analysis is then performed by means of a Fourier transform, whereby the whole data set is regarded as one single time series, including zeroes when data is not available. Figure 1 shows the resulting power spectrum in the five-minute range where spectral peaks are looked for. Having the solar result in mind, we are looking for a set of equidistant peaks representing the resolution of alternately even and odd degree eigenmodes. This power spectrum is evidently much noisier than the corresponding solar one (Grec et al, 1983). However, a regular pattern of about

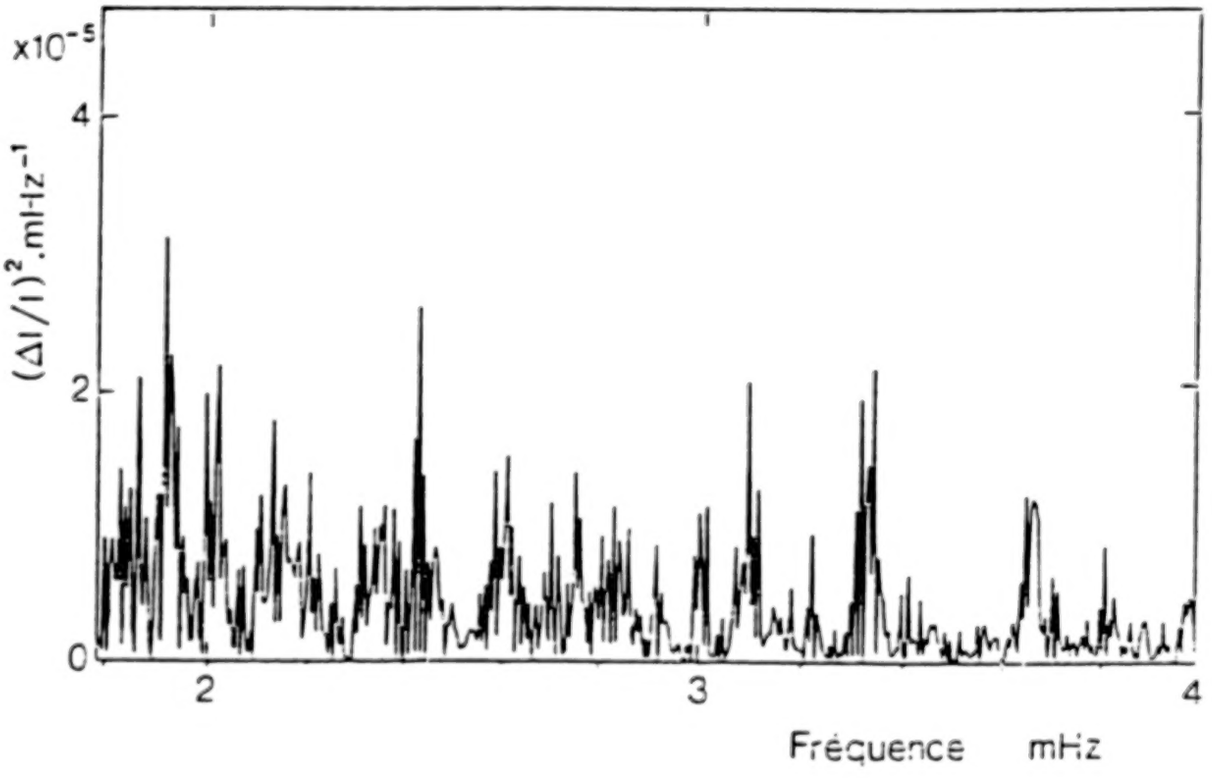


Fig. 1: *Part of the power spectrum of the data consisting in the monochromatic intensity in the red wings of the Alpha Centauri A Na D lines, recorded during three consecutive nights (May 1983) at the Cassegrain focus of the ESO 3.6 m telescope.*

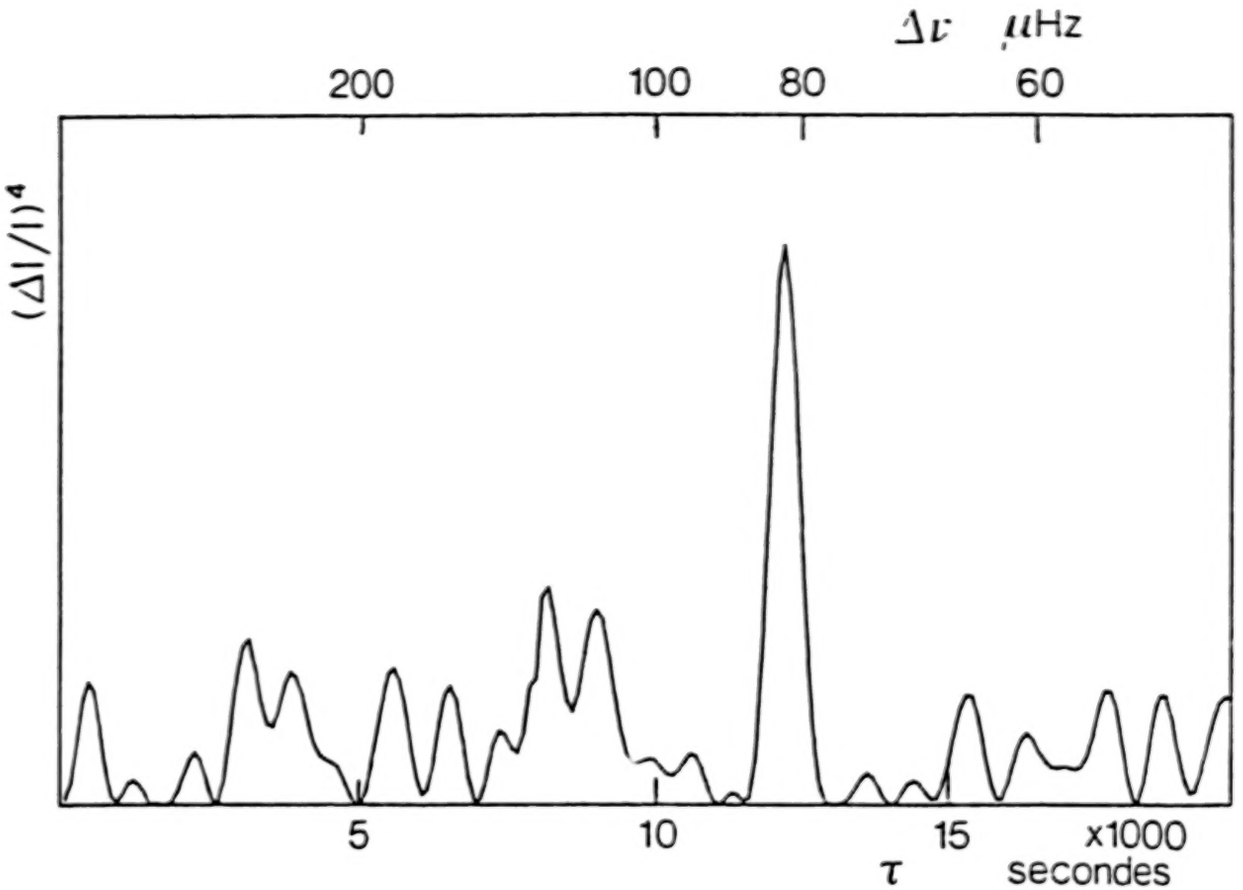


Fig. 2: *Power spectrum of the power spectrum limited to the frequency range 2.3–3.85 mHz, corrected of the square of the window function autocorrelation. The major peak, at 81.3 μHz, means that the expected periodicity in the signal spectrum is indeed present.*

80 μHz seems to be present just around 3 mHz. In order to check the significance of this possible pattern, the next step consists in looking for a periodicity by calculating the power spectrum of a given section of this power spectrum. This has the dimension of the square of an autocorrelation and therefore, the result shown in the Figure 2 has been corrected of the square of the autocorrelation of the temporal observing window. It shows convincingly that only one periodicity is present in the range 2.3 - 3.8 mHz of the power spectrum, with a period that happens to be 81.3 μHz.

Although significantly different, this result is of the same order of magnitude as the 68 μHz obtained in the solar case. It is then to be regarded as a very convincing evidence for the detection of five-minute p modes on Alpha Centauri.

Now, once admitted the existence of a pattern of equidistant peaks in the power spectrum, the data analysis can be pursued one step further by trying to extract this pattern from noise. This is done by using the knowledge of the periodicity (81.3 μHz) and phase of this period provided by the Fourier analysis whose result

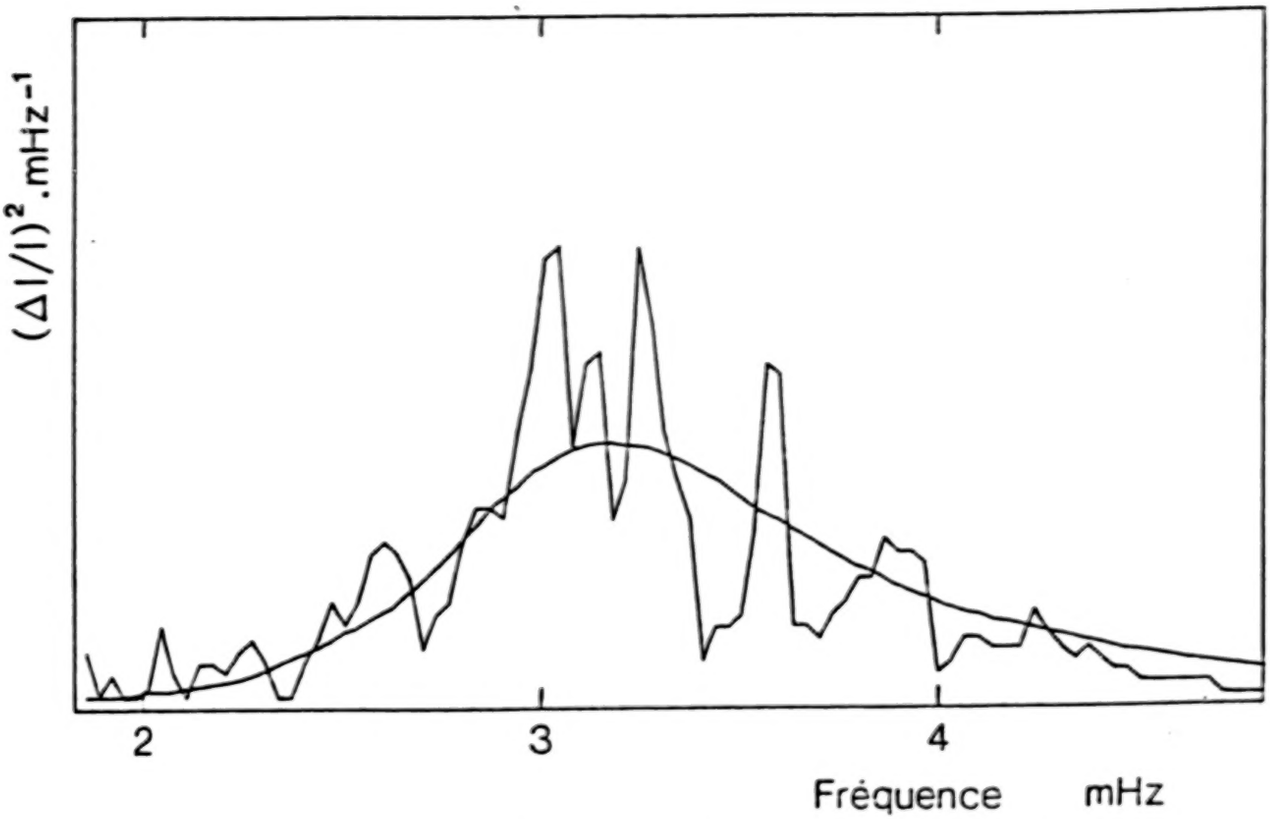


Fig. 3: *Using an adapted filtering, locked in frequency and phase, it is possible to extract from surrounding noise the power contained in the discrete pattern which is present in the power spectrum of Fig. 1 (black part). The continuous line shows, for comparison, the same envelope measured with an identically resolved solar power spectrum.*

is displayed in the Figure 2. An adapted filtering with this period and this phase is made on the power spectrum of the Figure 1 and with a resolution of 0.32 mHz (sinus fitting at locked phase on 0.32 mHz wide slices of the power spectrum). The result, shown in the Figure 3, is compared to the envelope of the solar spectrum obtained with the same resolution (from Fossat et al, 1978). The similarity is really striking and does not leave much room for doubt about the significance of the result obtained on Alpha Centauri.

Theoretical implications of this result must now be investigated. At first order, it seems that we are facing a serious problem. Indeed, the mass of the star is precisely known (10% more than the Sun), as well as its temperature (indistinguishable from the solar one). Although its southern location does not provide a parallax as accurate as we would like (just because it has been much less observed than many northern stars), the radius can then be estimated to be about 30% more than the solar one. Consequently, the travel time of a sound wave between center and surface has to be longer than it is in the Sun, and the solar 68 μHz frequency splitting has to be shifted towards a lower value, instead of the presently measured 81 μHz. More data is urgently needed in order to confirm this splitting and

to distinguish individual eigenmodes. Just four consecutive clear nights instead of three should make this possible.

An earlier and very similar version of this paper has been published in the ESO journal "The Messenger", by E. Fossat, G. Grec, B. Gelly and Y. Decanini.

References

Fossat, E., Grec, G. and Slaughter, C.: 1977, Astron. Astrophys. 60, 151

Fossat, E., Decanini, Y. and Grec, G.: 1983, The Messenger, 33, 29

Fossat, E., Grec, G., Gelly, B. and Decanini, Y.: 1984, The Messenger, 36, 20

Grec, G., Fossat, E. and Pomerantz, M.: 1983, Solar Phys. 82, 55

HIGH PRECISION MEASUREMENT OF STELLAR RADIAL VELOCITY VARIATIONS

William D. Cochran
The University of Texas at Austin

Abstract

We are building a prototype instrument for measurement of stellar radial velocity variations to a precision of a few meters per second. The instrument will be used to study low amplitude stellar non-radial oscillations, to search for binary systems with large mass ratios, and ultimately to search for extrasolar planetary systems. The instrument uses a stable Fabry-Perot etalon, in reflection, to impose a set of fixed reference absorption lines on the stellar spectrum before it enters the coude spectrograph of the McDonald Observatory 2.7-m telescope. The spectrum is recorded on the Octicon detector, which consists of eight Reticon arrays placed end to end. Radial velocity variations of the star are detected by measuring the shift of the stellar lines with respect the artificial Fabry-Perot lines, and correcting for the known motions in the solar system.

Introduction

Improvement of the precision in measurement of stellar radial velocity variations will have significant impact on several diverse fields of astronomy. The newly developing field of stellar seismology will benefit greatly form any increase in measurement precision. As measurement sensitivity increases, stellar oscillations become detectable in lower mass and lower luminosity class stars. These non-radial p-mode oscillations provide an important new probe of stellar internal structure. Detection of binary stellar systems through the barycentric orbital motion of the central star is an exciting goal of any radial velocity variation monitoring program. This is an exceedingly difficult task, and any such observing program will undoubtedly discover a very large number of new binary stellar systems with large mass ratios. These systems will help define the low end of the stellar mass function - objects between faint M dwarfs ($0.2M_{\odot}$) and massive planets ($0.001M_{\odot}$).

We are currently building a prototype instrument for McDonald Observatory to measure stellar radial velocity variations with a precision of a few meters per second. The instrument will use the McDonald 2.7-m telescope, the coudé spectrograph, and the new Octicon detector. The instrument itself consists of a stable Fabry-Perot etalon, used in reflection, to impose a set of fixed reference absorption lines on the stellar spectrum before it enters the spectrograph. Doppler shifts of the stellar absorption lines with respect to the Fabry-Perot artificial absorption lines are measured by a cross correlation technique.

Scientific Goals

Planet Detection: The original motivation for starting this instrumentation program was to search for planetary systems around other stars. The precision requirements for this task have been discussed extensively, (see e.g. Black 1980) and will only be summarized here. The velocity of the sun around the solar system barycenter, measured in the ecliptic plane, is about 13 meters/second with a period of 11.86 years. Detection of such a system requires an instrument precision of a few meters per second to be maintained for over a decade. The constraint of long term stability for an instrument is at least as difficult as achievement of the radial velocity precision. The observed velocity increases and the orbital period decreases as either the planetary mass increases or as the star-planet separation decreases.

Binary Stars: This technique, of course, is not limited to the problem of detection of planetary systems. The only parameter that really distinguishes a planetary system from a multiple star system is the mass of the secondary object. If the companion is more massive than the threshold at which fusion processes occur, it is a secondary star; otherwise it is a planet. This raises the question of whether giant gas planets as we know them (Jupiter-like objects of $\sim 0.001 M_{\odot}$) are formed by the same process as binary stars, or whether planet formation is a completely different mechanism. At present, we have virtually no information on the objects covering the two orders of magnitude of mass between Jupiter and the faint red dwarf stars. Objects which may be in this mass range are now being discovered in astrometric studies of binary systems (Harrington 1983), but progress is slow and difficult. A radial velocity search of bright stars for low luminosity companions would fill in much of this mass range. The shape of the low end of the stellar mass function is largely unknown. While it is unlikely that a large amount of mass can be "hidden" in low luminosity secondary stars of binary systems, it is very possible that a very large number of such stars exist. When Abt and Levy (1976) attempted to correct their study of multiplicity in F2-G3 IV and V stars for their detection threshold (~500 m s^{-1}) and selection effects, they concluded that there should be no solitary field solar type dwarfs. If this is indeed true, it is telling us something profound about the process of star formation. Obviously, it is exceedingly important to survey field solar type dwarfs with a hundred-fold increase in precision over the Abt and Levy study.

Stellar Oscillations: The new field of stellar seismology has burgeoned with the increase in precision of radial velocity variation measurement. Helioseismology has provided exciting new insights into the internal structure and composition of the sun (see e.g., Bonnet 1983, Deubner 1983). It is quite possible that nearly all stars with sizable convective zones may be somewhat unstable to stochastically excited non-radial oscillations similar to the solar 5-minute oscillations. Detection of these oscillations in other stars will be extremely difficult since the amplitudes are low and the disk cannot be resolved. Non-radial modes of high degree ℓ will be rendered undetectable in other stars since the positive radial velocity contribution of one element of the stellar surface will be largely canceled by the negative contribution of the adjacent surface. In general, disk integrated observations will be dominated by modes with $\ell \leq 6$. This problem, however, certainly does not limit the usefulness of such data. An oscillation of given degree ℓ samples a depth in the star given by

$$d_{min} = (2n+3)R/\ell.$$

Thus the low degree modes, the oscillation which are most easily detected in disk integrated data, sample virtually the entire star. If multiple modes are present, then measurement of the frequency separation, $\Delta\nu$, is essentially a measurement of the radius of the star (Christensen-Dalsgaard and Frandsen, 1983).

Radial Velocity Detection

Classical techniques for measurement of stellar radial velocities using a separate comparison source spectrum rarely achieve precision better than a few tenths of a kilometer per second. Griffin and Griffin (1973) showed that the limiting factor is the difference in illumination of the spectrograph optics by the stellar and comparison sources. The stellar beam traverses the terrestrial atmosphere and telescope optics, and is thus subject to seeing and guiding fluctuations. The comparison spectrum is taken at a different time, has a different beam illumination pattern, and often follows a different geometrical path through the spectrograph.

The Griffins realized that these problems could be solved if wavelength comparison lines could be imposed on the stellar spectrum before the light enters the spectrograph. They suggested the use of telluric absorption bands as the calibration lines, an idea that had been used nearly a century earlier (in 1877) by Duner (1906) to measure the solar rotation. In this method, both the stellar and reference spectrum traverse the same optical path. The method, obviously, may be used only in a spectral region with suitable telluric bands, such as the 6300Å O_2 band. For the method to work, the stellar lines can not be too shallow or broad, the exposure must be reasonably short, and the spectrograph must be exceptionally stable. The resulting radial velocity precision is limited by the variability of the telluric lines resulting from winds, and variations in temperature and pressure. The

Griffins concluded that precision of about 10 m s^{-1} could be achieved using suitable telluric bands and photographic bands. Application of this technique by Young et al. (1979) to spectra of Venus resulted in an internal error for individual plates of ±16 m s^{-1}. M. Smith (1982, 1983) has applied this technique to time series observations of Arcturus and Aldebaran. He used the McDonald Observatory 2.7-m coudé spectrograph and a single unintensified Reticon Corporation RL1024B detector. He measured positions of stellar and telluric lines by calculating a least squares fit of a polynomial function to the line profile. He claimed an rms relative velocity error of ±6-8 m s^{-1} on good nights. He concluded that this method was limited in precision by the variations in the telluric atmospheric conditions, rather than by the stability of the spectrograph or the detector.

To improve the precision of this technique, one must improve the stability of the reference line system. Instead of using the telluric atmosphere to provide the reference absorption lines, Campbell and Walker (1979) reasoned that stability could be improved with a gas absorption cell in front of the spectrograph, They could easily control the temperature and pressure of the gas in the cell. The gas must have a sufficiently strong absorption band in the accessible spectrum (3000A < λ < 11000A). The cell must be manageable in size, and the gas must be either diatomic or linear polyatomic so that its spectrum will be simple and clean. The B value must be several cm^{-1} so that its lines are well separated and unblended. The best choice is the HF 3-0 R branch near 8700 A. Campbell and Walker achieved an rms deviation of ±15 m s^{-1} on their first system at the Dominion Astrophysical Observatory. A second system at the CFHT coudé spectrograph, using a single Reticon RL1872F detector, is also capable of ~15 m s^{-1} precision (Campbell et al., 1981).

When we evaluated these efforts by others, we realized that we already had the key ingredients of a very high precision radial velocity spectrometer. Since M. Smith's results, which are among the highest precision data reported by anyone, were limited by the stability of the reference lines, then the intrinsic stability of the spectrograph and detector must be better than the values he reported. This system is currently being improved with the installation of the Octicon. This detector consists of eight Reticon arrays, each with 1872 silicon photo-diodes, placed end-to-end. Use of the Octicon detector increases the number of discrete data points in a spectrum from the 1024 of M. Smith's data to nearly 15000. Campbell et al. (1981) achieved a measurement precision of ≤0.003 pixels using the same type of Reticon chip as we are using in the Octicon. If we can develop a reference line system at least as stable as Campbell's, then the achievement of 0.003 pixel precision with our instrument corresponds to 0.198mA, or 8.5 m s^{-1} at 7000A per Reticon. The multiplexing of the eight Reticons in the Octicon should give a final precision of ≤3.0 m s^{-1}. An improvement in the reference line system should make higher precision attainable. Thus, we already had everything we needed except for the ideal reference line system.

The McDonald Instrument

The perfect system for generating reference lines would allow us to place strong, narrow, absolutely fixed reference absorption lines anywhere we wish within the stellar spectrum. The "perfect" machine does not exist, but an excellent device is a Fabry-Perot etalon, used in reflection. The Fabry-Perot will give us a series of deep artificial absorption lines corresponding to the interference maxima of the etalon. They will be spaced regularly throughout the entire spectrum. We have the freedom of selecting a region of the stellar spectrum with a reasonable density of stellar lines, but also contains a large amount of continuum which is free of weak stellar lines, to avoid excessive blending of Fabry-Perot lines and stellar lines.

Our entire radial velocity measurement system consists of the McDonald Observatory 2.7-m telescope, our prototype Fabry-Perot interferometer assembly, the coudé spectrograph, and the Opticon detector. The coudé spectrograph, when used with any of three 1200 groove/mm gratings, gives a reciprocal dispersion of 4.4 Å/mm at the focus of the six foot camera where the Opticon is being installed. The Opticon records simultaneously eight portions of the spectrum of a star, each covering 123Å at 0.132Å resolution. There is a 71Å gap between each of the spectra which can be filled, if desired, by a second exposure at a slight grating rotation.

The design of our prototype Fabry-Perot interferometer assembly is shown in Figure 1. The interferometer is housed in a sealed, temperature controlled unit which is placed in front of the entrance slit of the 2.7-m coudé spectrograph. The instrument baseplate is mounted on an optical bench which is part of the spectrograph foundation and is mechanically isolated from the building and dome.

The artificial reference absorption lines are generated by reflecting the incoming starlight off of the Fabry-Perot etalon. Absorption lines will then occur at the interference maxima of the etalon, given by:

$$m \lambda = 2 n d \cos(\theta)$$

where m is the interference order number, λ is the wavelength, n is the index of refraction between the etalon plates, d is the spacing of the etalon gap, and θ is the angle of incidence. The Fabry-Perot etalon is the heart of the instrument. The ultimate precision we achieve is dependent on our ability to construct, mount, and maintain the etalon so that the interference orders will not move. We will use a conventional fixed "air-spaced" etalon, with the plate separation maintained by three optically contacted precision spacer buttons. We choose not to use conventional plate adjustment techniques, such as application of pressure with nylon tipped screws, since such adjustments are very difficult to maintain constant for a multi-year time period. We wish, instead, to use a precision etalon that does not require any adjustment. The entire optical assembly is kept in an enclosure with temperature regulated to better than ±0.1°C.

It is obviously advantageous to keep the spread of incidence angles, $\delta\theta$, as small as possible. This is done, to first order, by collimation of the beam. However, even within a collimated section there is a small $\delta\theta$ resulting from the finite size of the entrance aperture. The effects of this unavoidably small spread of the collimated beam may be minimized by using the etalon as near to normal incidence as possible. To use the etalon at exactly $\theta = 0$, we would need to physically separate the beams using an optical isolator system (Cochran, Smith, and Smith 1982). This type of system, however, introduces polarization to the beam, and requires two matched etalons and a large number of high precision transmitting optical components. If we are able to tolerate illumination of the Fabry-Perot at a slight off-axis angle, then we may use any one of a number of standard collimator-camera designs normally used for spectrographs. In Figure 1 we show an astigmatism compensated Ebert-Fastie system. As is well known, a Czerny-Turner system in which the collimator and camera mirrors are each part of a single spherical mirror results in complete compensation for tangential coma (Fastie 1952a), but gives serious problems with astigmatism. The standard means of correction for astigmatism with curved entrance and exit slits (Fastie 1952b) will not work in our application, because we must feed the coudé spectrograph with a straight slit image. An ingenious method for compensating for the astigmatism by use of convex spherical folding mirrors was given by Rosendahl (1961; 1962a; 1962b). The resulting system leaves only a tolerably small amount of uncorrected spherical aberration.

The Ebert-Fastie system requires the collimated beam to strike the etalon at about $\theta = 5°$. We are thus using the etalon in part of its off-axis ring pattern rather than in the central axial fringe. In so doing, we must make sure that the radial extent of the entrance aperture does not unduly degrade the finesse of the Fabry-Perot system. Our calculations show that this is no problem as long as the entrance aperture is a long thin spectrograph slit rather than a circular star image.

Calibration and Data Reduction

The ability to obtain stellar spectra with our instrument is, of course, only part of the overall problem. We must also develop a method of reducing the data and determining the radial velocity shift of the star. We emphasize again that we are not attempting to make <u>absolute</u> radial velocity measurements. Rather, we wish to measure relative <u>variations</u> of the radial velocity of the star with respect to the center of mass of the solar system. Therefore, our data reduction task consists of three basic parts: 1) determination of the Doppler shift of the stellar spectrum, 2) calculation of the motion of McDonald Observatory around the solar system barycenter, and 3) determination of any possible motions, drifts, or shifts within the Fabry-Perot instrument.

Each of the eight independent spectra from the Octicon consists of both the stellar lines and the artificial reference lines imposed by the Fabry-Perot etalon. Figure 2 shows what typical data will look like. The upper

spectrum is the solar spectrum near 6420A taken with a single 1872 diode Reticon at the same dispersion the Octicon will use. The second spectrum is a calculation of the artificial absorption lines which will be generated by the etalon. The bottom plot is the product of these two, or an example of what will be detected by each of the eight arrays of the Octicon. We will measure our relative velocities with respect to a high signal/noise "standard" spectrum of each star taken at the beginning of the program. Measurement of these shifts will be done with a cross-correlation technique separately applied to each of the eight Octicon spectra. There are two different line shifts we must measure with the cross correlations. The first is the shift in position (Reticon diode number) of the reference lines of the program spectrum with respect to the reference lines of the standard spectrum. Such shifts result from any movement (grating rotation, temperature changes, etc.) within the spectrograph-detector system, and will be present in both the Fabry-Perot reference lines and the stellar spectrum. The second shift is the shift of the stellar absorption lines in the program spectrum with respect to the stellar lines in the "standard" spectrum. The difference between these two shifts reflects the change in radial velocity of the star with respect to the McDonald Observatory 2.7-m telescope.

Correction of this shift for the motion of the telescope around the solar system barycenter is straightforward. All of the relevant motions within the solar system are quite well known. In fact, the McDonald Observatory lunar laser ranging program, which has operated routinely from the coudé focus of the same McDonald 2.7-m telescope since 1969, has determined the lunar orbit, the rotation of the earth, and the location of McDonald Observatory, each to a precision well beyond our needs for this program. The result of the correction of the calculated shift for these known motions gives us the final datum we desire, the change in radial velocity since the program started.

We have taken every precaution to insure that the artificial absorption lines we impose on the spectrum will not move with time. However, we need a method to calibrate the interferometer so that we can detect and account for any motion which might occur. One method of calibration involves the use of a very convenient set of fixed reference lines. Since our spectra will be centered near 7000 A, the O_2 "A" band (≈7600A) is within the region for which the optical coatings are still good. The absorption lines produced by the oxygen within the spectrograph are strong enough to be easily measured. Cross-correlation calculations of the shift of the coudé room oxygen lines with respect to the Fabry-Perot lines will provide an independent check on the stability of the interferometer after the O_2 line positions have been corrected for changes in the spectrograph temperature and pressure. The solar spectrum may also be used as a calibration source of somewhat lower accuracy by observation of a known fixed location on the lunar disk. The observed shifts of the solar lines with respect to the Fabry-Perot orders must be corrected for the motion of the Observatory with respect to the sun and the moon, which are quite well known. However, use of the solar spectrum as a radial velocity standard has the uncertainty of apparent shifts due to sunspots rotating across the solar disk, and of low amplitude solar oscillations.

I would like to thank Dr. Wm Hayden Smith, Dr. Harlan J. Smith, and Dr. Brenda Young for their valuable assistance and advice. This work has been supported in part by a grant from the Research Corporation, by NSF Grant AST8315379, and by NASA Grant NGR 44-012-152.

REFERENCES

Abt, H.A., and Levy, S.G. 1976, Ap. J. Suppl., 30, 273.
Black,D.C. 1980, Space Science Rev., 25, 35.
Bonnet, R.M. 1983, Solar Physics, 82, 487.
Campbell, B. and Walker, G. A. H. 1979, Pub. A.S.P., 91, 540.
Campbell, B., Walker, G.A.H., Johnson, R., Lester, T., Yang, S., and Auman, J. 1981, Proc. SPIE, 290, 215.
Christensen-Dalsgaard,, J. and Frandsen, S. 1983, Solar Physics, 82, 469.
Cochran, W.D., Smith, H.J., and Smith, W.H. 1982, Proc. SPIE, 331, 315.
Deubner, F.-L. 1983, Solar Physics, 82, 103.
Duner 1906, Uber die Rotation der Sonne (Upsala:).
Fastie, W.G. 1952a, J. Opt. Soc. Am., 42, 641.
Fastie, W.G. 1952b, J. Opt. Soc. Am., 42, 647.
Griffin, R. and Griffin, R. 1973, M.N.R.A.S., 162, 243.
Harrington, R. 1983, Bull. AAS, 15, 799.
Rosendahl, G.R. 1961, J. Opt. Soc. Am., 51, 1.
Rosendahl, G.R. 1962a, J. Opt. Soc. Am., 52, 408.
Rosendahl, G.R. 1962b, J. Opt. Soc. Am., 52, 412.
Smith, M. A. 1982, Ap. J., 253, 727.
Smith, M.A. 1983, Ap. J., 265, 325.
Young, A. T., Schorn, R. A., Young, L. D. G., Crisp, D. and Barker, E. S. 1979, Icarus, 38, 435.

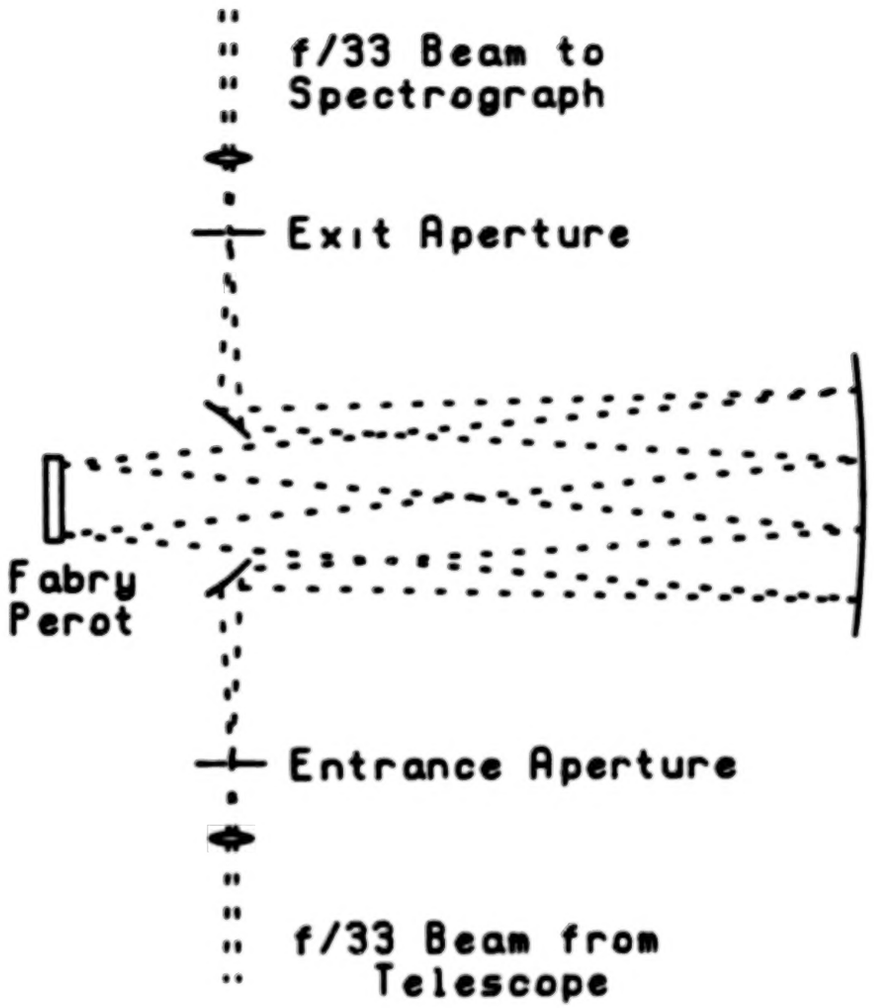


Figure 1. - A schematic diagram of the McDonald high precision radial velocity interferometer.

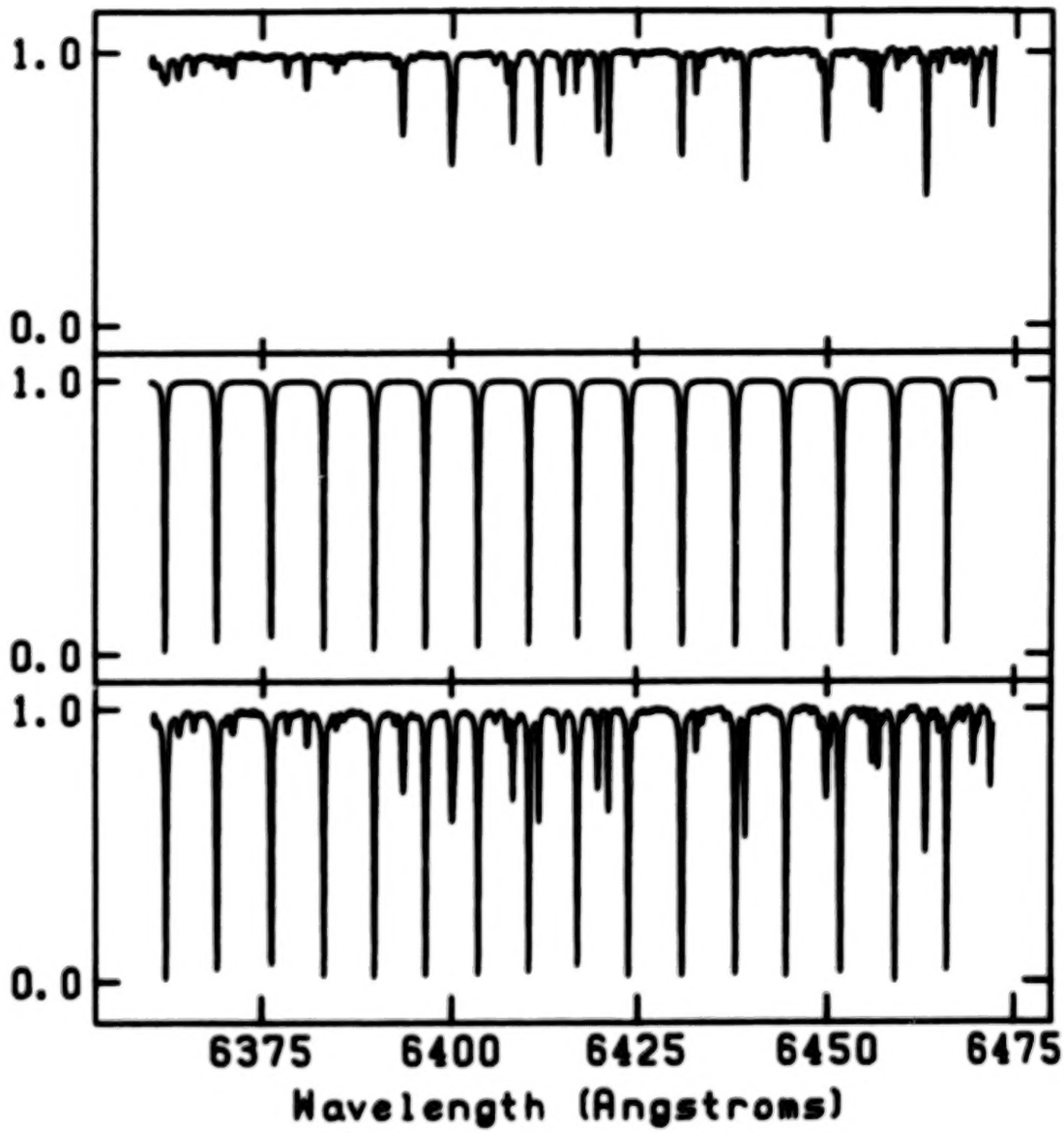


Figure 2. - Some sample data. Top spectrum - the solar spectrum near 6420A recorded on a 1872 diode Reticon detector at 0.132A resolution. Middle spectrum - the artificial reference absorption lines imposed by the Fabry Perot etalon. Bottom spectrum - The product of the above two spectra, as it will be recorded by the instrument.

N85-17897

A SURVEY FOR PHOTOMETRIC VARIABILITY FROM SPACE

Hugh S. Hudson
CASS C-011, UCSD
La Jolla, CA 92093 USA

ABSTRACT

A survey for photometric variability in a wide variety of astronomical objects would produce much new information about their interiors and dynamics. This paper discusses reasons for such a survey, gives the example of the solar-constant variations as a guide to what may be expected from main-sequence stars, and proposes a concept for a satellite dedicated to a survey of photometric variability.

1. INTRODUCTION

The variability of an astronomical object contains information about its internal structure. The time series of the flux variations at different wavelengths contains all of this information. We often summarize the variations by calculating the power spectrum, which can show lines (oscillations) and continua (random phenomena) just as the radiation spectrum can. Obtaining a complete set of data may be essential for characterizing the variability, but heretofore most astronomical photometry has necessarily generated time series with many gaps. The more variable an object, or equivalently the more sensitive the photometry, the more necessary it is to generate a complete time series.

Photometry as an astrophysical tool has not received the development effort of other branches of observation. Precise or accurate photometry of course is limited when ground-based observatories must be used, just as is almost anything else that an astronomer might wish to do. Operating photometric instruments from space is the obvious solution, but one that is so cumbersome and expensive that little has yet been done. The Hubble Space Telescope will provide excellent photometry and will allow us to test some of the systematic limits of space photometry, but it cannot be used for the tedious monitoring that is necessary for obtaining a complete time series on a given object.

There is one example of a complete and precise photometric time series from an astronomical object: the Sun. After many years of solar space astronomy, high-quality observations of the solar constant finally commenced in 1980 with the Solar Maximum Mission. We discuss these observations in section 3 below in terms of their significance for space photometry.

The main purpose of this paper is to present ideas for surveys for photometric variability, emphasizing space-borne instrumentation. There is no question that dramatic improvements in the quantity and quality of ground-based

photometry can be achieved, so that some of the reasons listed below for a photometric survey (section 2) will disappear with time. Equally, it is also clear that there are fundamental systematic limits at some level for ground-based work, and that the ultimate future of photometric astronomy is in space.

There are three conflicting requirements for a photometric variability survey:

> As many objects as possible should be surveyed. In fact, doing all of the stars below a given magnitude limit would be the best idea.
>
> Each object should be observed continuously. The entire spectrum of variability may be interesting, and a complicated observing pattern ("window function") with many gaps in it will reduce the accuracy of the spectral measurement, possibly in important frequency ranges (for stationary processes) or at important times (for unique variations).
>
> Each measurement should be as precise as possible. Great precision allows the measurement of subtle variations, such as the stellar normal-mode oscillations.

A photometric survey consists of measuring the brightness variations of a large number of objects. Ideally this would be done in via absolute photometry so that different objects could be intercompared. This is in practice difficult, so that ground-based photometrists rely upon intercomparisons among stars to perform relative photometry, either "all-sky" or differential (Young, 1974). The solar constant observations described below are an example of "direct" photometry, in which the calibration of the detector is assumed to remain constant; even if its absolute photometric calibration cannot be determined more accurately than about 0.1%, its relative readings on the same object (the Sun) give us a reliable time series with which to characterize the variations. This kind of photometry can best be carried out in space, where the instability of the atmosphere makes such direct photometry meaningless for variations on longer time scales.

2. WHY MAKE A VARIABILITY SURVEY?

What scientific problems can we tackle with direct stellar photometry from space? There appear to be five major categories:

> 1. Ordinary stellar variability at low amplitudes. Many known categories of variable star will have members exhibiting only small variations, which would be very interesting to detect. One example is spottedness at the solar level (< 0.3%; see Willson et al., 1981), rather than at the BY Dra level; Section 3 gives more details.
>
> 2. Astroseismology. Main-sequence stars may show photometric evidence for non-radial oscillations at low levels; the sun is an example (Woodard and Hudson, 1983). Other types of stars will have

larger-amplitude variations from a variety of causes (e.g. Kurtz, 1982). Astroseismology gives information on stellar interior structures and dynamics (Christensen-Dalsgaard and Frandsen, 1983; Gough, 1984).

3. Extragalactic variables. Active galactic nuclei vary for unknown reasons. Frequent monitoring and access to smaller variations are of obvious importance.

4. Transients. Many classes of singular variations exist: supernovae, novae, optical counterparts of gamma-ray bursts or the like, planetary occultation, flares on stars of many types. For these, a stable observing program in reproducible conditions will provide an ideal data base.

5. Solar-system Objects. The zodiacal light, coronal plasma clouds, comets and asteroids, all would be detectable by a photometric survey from space. These objects, their motions and time variations, are all interesting and difficult to observe from the ground.

While none of these individually may produce a perfect argument in favor of space rather than ground-based observations, together they are compelling. There are unique properties of astronomical variability that are accessible only to a space observatory, and which will require properties quite different from those of existing satellites or the Space Telescope. The photometric data base that could be produced by a satellite such as that sketched below would serve a very large community of astronomers, both observers and theorists, much in the same way that IUE has done.

In considering a program of precise space photometry, it is very important to realize that many of the objectives of the program can be met by ground-based photometry. For the most rapid progress, efforts in both directions should proceed simultaneously. An effective ground-based program is an excellent means for obtaining direct experience with new types of instrumentation as well as for complementary scientific work. Ideally, the very observers and theorists working with precise ground-based photometry will become the principal users of the satellite data.

3. FOUR COMPONENTS OF SOLAR VARIABILITY

Observations from the Solar Maximum Mission have given us a taste of what the variability spectrum of a star might contain. These measurements give the total solar irradiance (integrated over the whole spectrum and over the entire disk of the Sun) every two minutes. The probable error for a 96-minute integration is on the order of 15 parts per million. These data reveal the existence of (at least) four kinds of variation that could not have been detected from the ground. These are (i) the effects of active regions (Willson et al., 1981); (ii) a continuum of variations due to surface granulation (see Figure 1); (iii) the "p-mode" global oscillations (Woodard and Hudson, 1983), and (iv) the steady trend detected over the interval 1980-1984 (Willson, 1984). Figures 1-3

show various aspects of the ACRIM data.

Table 1 gives a rough idea about the precision and sampling pattern needed to characterize the four identified components in a star similar to the Sun. The detector stability or calibration reference must be good on the time scale of the variation - tens of minutes for solar-type p-mode oscillations; tens of years for solar-type magnetic cycles. The solar observations are relatively new and much analysis is under way, but it is clear that each of the four components thus far identified contains some information about solar interior structure or dynamics. An extrapolation of this experience to other stars will be the key to gaining observational knowledge of the physical processes in stellar interiors.

Table 1
Observing Parameters for
Four Solar-type Variations

Variation Type	Sampling Interval	Sampling Duration	Precision per point
Secular Trend	One hour	Many years	300 ppm
Active Regions	One hour	Many years	300 ppm
Granulation	10 minutes	Few days	100 ppm
Oscillations	One minute	One month	100 ppm

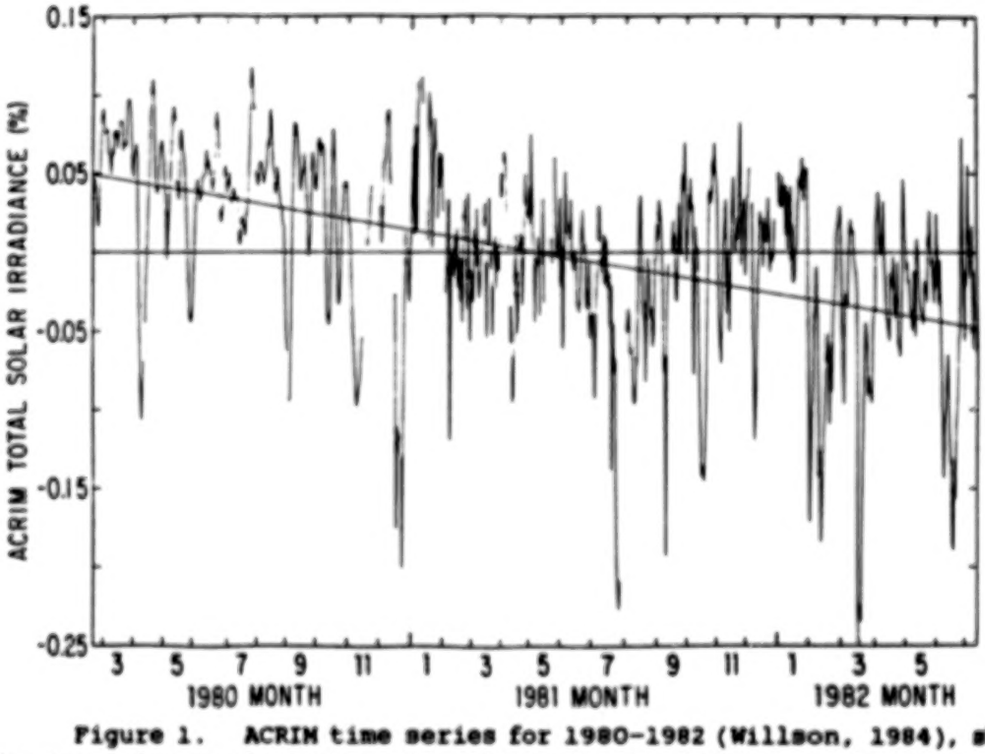


Figure 1. ACRIM time series for 1980-1982 (Willson, 1984), showing the trend downwards during this interval. The large negative excursions, on the order of 0.1 - 0.2 %, result from sunspot passage across the solar disk.

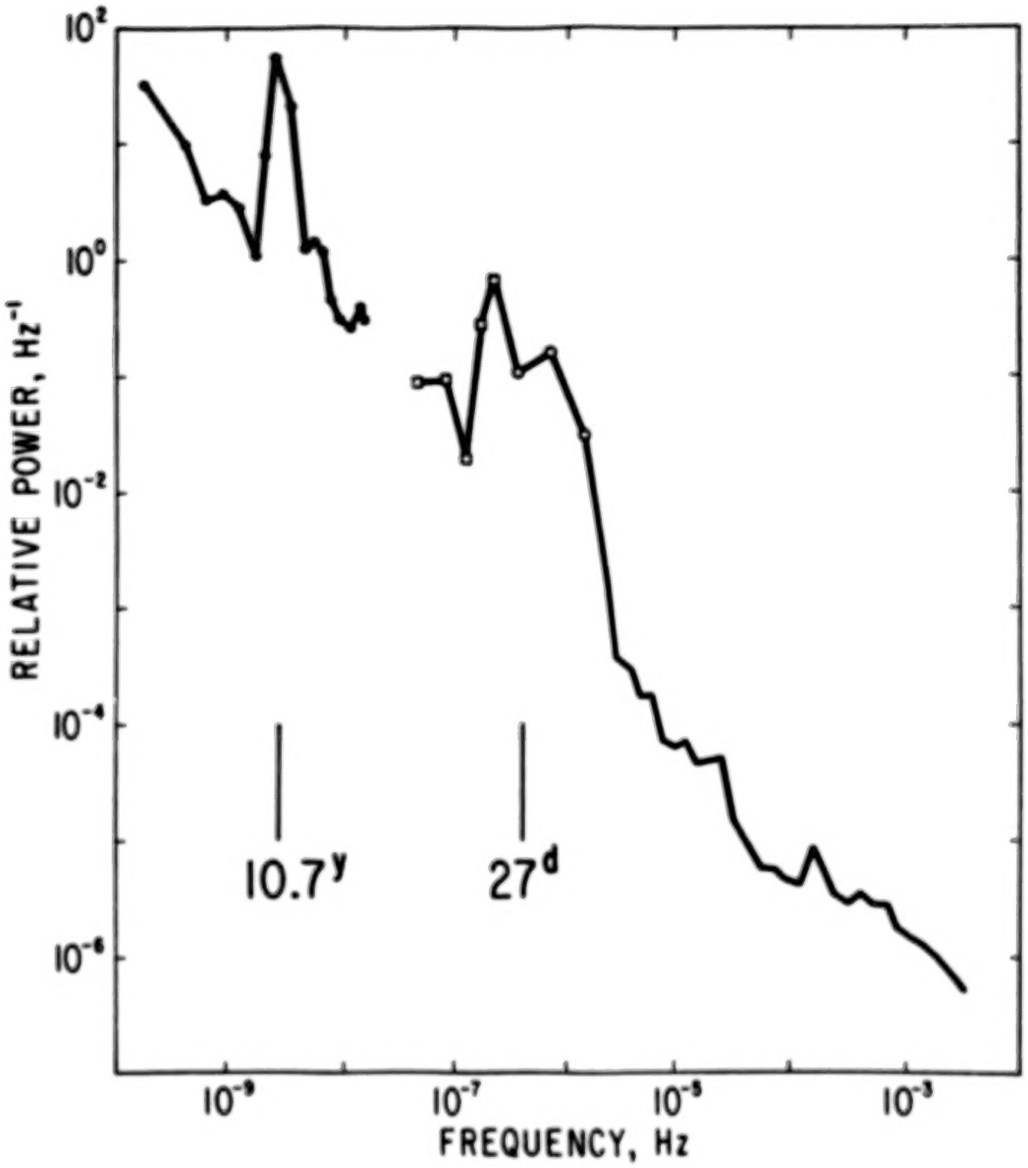


Figure 2. Power spectrum representing the solar irradiance variations observed by the ACRIM instrument in 1980 (Woodard et al., 1982). The key features are the broad peak in the vicinity of the solar rotation period (individual points of the periodogram are plotted at the lowest frequencies), and the continuum of variations at higher frequencies (spectral estimates re-binned into 20% frequency increments).

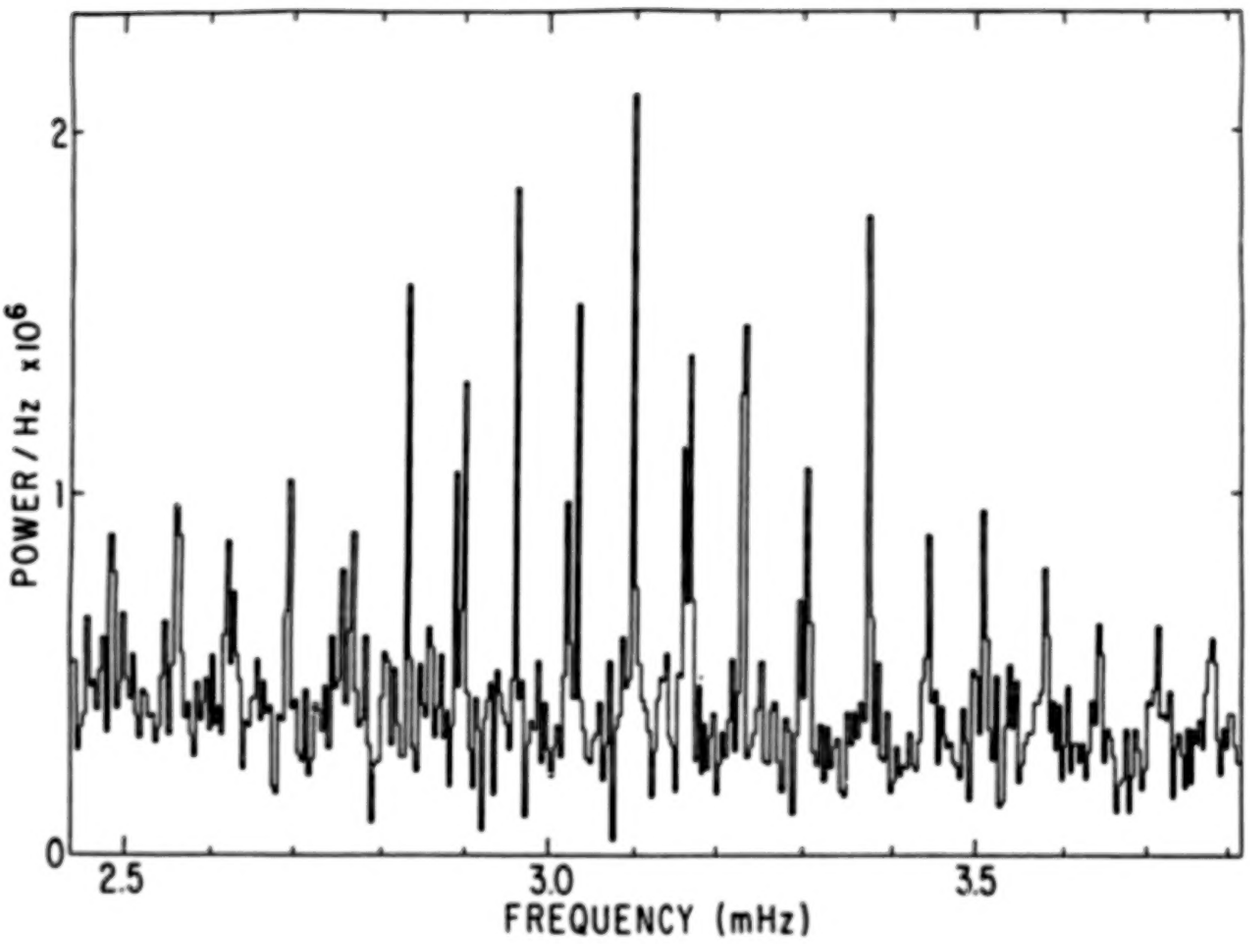


Figure 3. High-frequency power spectrum from the same ACRIM time series, showing the p-mode "5-minute" oscillations of degree ℓ = 0, 1, and 2 (Woodard and Hudson, 1983).

4. CONCEPT OF A SCANNING PHOTOMETRIC SATELLITE

This section describes one scheme (the "All-Sky Photometric Explorer"; Hudson, 1984) for carrying out a survey. Table 2 summarizes the key points of the concept.

The satellite instrumentation would consist of a number of photometric imaging cameras based upon CTD (Charge Transfer Device) sensors. A "photometric imaging camera" is meant to be a device for which the entire focal plane can be used effectively for direct photometry. This means that non-uniformities not only of detector response, but also of the entire optical system, must be understood quantitatively and corrected in the frequency range of the interesting variations. An ability to do this is essential to this concept of a sensitive variability survey of a large number of objects. The CTD could be either a CCD (Charge-Coupled Device) or CID (Charge-Injection Device); there is considerably more experience in astronomy with the former but both types have advantages.

The satellite is assumed to be spin-stabilized for cheapness, and in geosynchronous orbit to minimize interruptions in the coverage of an individual object. This means that the cameras will scan the sky as the satellite rotates. A camera pointed radially outwards from the spin axis will scan a swath of sky defined by its field of view, on a great circle. Cameras pointed axially will scan small circles at a given point on the image plane of the camera. To establish photometric information, sophisticated on-board data manipulation would probably be required to permit individual objects to be tracked and measured.

Real-time data manipulations on board the spacecraft also are a requirement for the severe data compression needed to condense the information of many images into the brightnesses and positions of individual stars. On-board data reduction would be necessary to reduce the the telemetry requirement to a feasible level. The ground-based model for on-line data reductions of this type is the Steward Observatory transit telescope (McGraw *et al.*, 1982), in which the sky survey must be reduced to catalog form as the data accumulate.

Table 2
All-sky Photometric Explorer

Stabilization	Slowly spinning (T ‹ one hour)
Orbit	Geosynchronous
Detectors	CTD cameras (~ 12 total)
Telemetry	Photometric Reductions

5. CAMERAS

The individual cameras would have identical sensors but different optics and filtering, as well as different data treatment. Some cameras would be operated in a "spin-scan" mode (Mackay, 1982; McGraw *et al.*, 1982); in this kind of

operation the CCD transfers charge at a rate that matches the motion of the stellar images, so that charge from a given object builds up continuously until those pixels reach the edge of the detector. Other cameras might be operated in a more standard scan of individual frames, with integration times chosen to place the dynamic range of the sensor in an appropriate magnitude range.

A camera with a large field of view would be useful for observation of bright stars, for example, since confusion of stellar images would degrade its observations of faint objects. Such a camera would not be operable in the spin-scan mode since the long scan times on bright stars would simply produce saturation; instead it should produce a sequence of images for which on-board processing might be required to link together before reduction to catalog format.

Narrow-field radial cameras, on the other hand, could observe faint objects in the spin-scan mode and should be designed to reach magnitude levels suitable for extragalactic objects such as active galactic nuclei.

The axial cameras would observe stars as they moved in circles in the image plane. Such cameras would require a frame-scan readout; they would provide long looks at relatively faint stars and would be most suitable for "astroseismology."

Finally, all-sky cameras with very large fields of view would be useful in monitoring for the occurrence of bright transient phenomena. Transient-recognition logic would be needed (see Ricker et al., 1984, for a discussion of a comparable ground-based program) and this detection mode would be interesting for other cameras as well. Such cameras would probably use very-wide-field "fisheye" optics, possibly convex mirrors patterned on the auroral all-sky camera.

6. PHOTOMETRY WITH CCD's

Photometric imaging with CCD cameras presents special problems not found in traditional photometry. The image of the object falls on a group of pixels from which the signal must be extracted, rather than into an otherwise empty circular aperture defining the solid angle of integration. The pixels are rectangular and will have individual properties (the "flat field" correction) that require special treatment. On the other hand the creation of an image gives much stronger control over uncertainty in background signal and for rejection of nearby confusing objects. King (1983) has discussed the precision of photometry and astrometry with digital detectors as a part of the preparation for Space Telescope observations (e.g. Westphal, 1982). At the level of precision needed for ground-based observations, CCD detectors appear to be as reliable photometrically as photomultiplier tubes (Walker, 1984).

There is a long list of technical problems associated with CCD's. Among these we know of pattern noise, non-linearity in response, charge leakage during transfer, saturation and the likelihood of streaks of dead pixels resulting from it, sensitivity to cosmic rays, and interference effects in the thin layers of the sensor. No doubt others exist. All of these, together with the more usual problems of photometry (e.g. Young, 1974) will affect the ability of the CCD cameras to do relative photometry (the comparison of one group of pixels with another).

The most dangerous two effects for direct photometry (the monitoring of one object to generate a time series) seem at the present time to be (i) the

stability with time of the flat-field correction, and (ii) the residual uncertainty in the flat-field correction that couples scanning motion into apparent time variations. The former problem will essentially degrade the long-term stability of direct photometry, while the latter will affect the short-term stability by introducing noise on time scales comparable to the pixel-crossing time scale.

For a one-hour spin period, the pixel crossing time for an axial camera viewing an object 1.5 degrees off-axis would be about two seconds (assuming a 500x500-element CCD). Noise at the corresponding frequency of 500 mHz would be a consequence; interesting periods for seismology studies of main-sequence stars are lower, so that direct pixel-to-pixel variations would not present a problem. Of course, noise on all spatial scales will exist on a given CCD chip, and the extent to which this noise exists, can be characterized, and undergoes time variations remain to be determined in the laboratory and in space.

The problem of calibrations has not been discussed. In general, the philosophy assumed here is that the mass of data accumulated in the actual observations will serve, through statistical analyses, as an effective tool for self-calibration. This assumption may be wrong!

7. ALTERNATIVE SCHEMES AND OTHER MISCELLANEOUS IDEAS

(a) "Planetarium Camera": Alternative schemes using three-axis stabilization could make precise photometry much easier. One such scheme would put an individual detector with its own field lens at the correct position in the focal plane for a given star; the telescope would stare at the same direction permanently to provide a photometric time series for each of the stars for which detectors were provided. This scheme could be made flexible by moving the detectors at intervals, perhaps through fiber-optics feeds.

(b) Despun platform: A despun platform would not be so expensive as a three-axis stabilization. The prototype for such a spacecraft might be the OSO type; for all-sky photometry the spinning section could carry the survey cameras, while the pointed section could carry the "seismology" cameras. "OSO" could now stand for Orbiting Stellar Observatory!

(c) Multiple detectors: Camera designs with multiple CCD's in the focal planes offer several advantages. One defect of the CCD sensors is their small size, which limits the aperture of the camera. Alternative methods of obtaining larger effective detector areas should be explored. A separate advantage derives from the possibility of multiple scans of the same field, helping to remove noise sources such as cosmic rays.

(d) Dichroic filters: Individual cameras with multiple color capability would allow a reduction in the number of cameras as well as a some new science. For example, monitoring the K-index would allow studies of chromospheric activity in stars and the observation of stellar flares even in solar-type stars (Livingston and Ye, 1982). Dichroic filters would not be suitable for the cameras dedicated to the highest precision, because of additional non-uniformities they would introduce (Young, private communication).

(e) Other types of detector: Alternatives to CTD detectors are certainly

worth considering. Detectors with larger formats would be especially interesting, since in this concept a larger detector would translate into a greater duty cycle or fraction of sky coverage. On the other hand, high quantum efficiency and capability for handling high photon arrival rates are also important, since many kinds of faint variability would be limited by photon statistics.

(f) Diffuse background sources: Sky brightness - light pollution or airglow - are not a problem in space, but other nuisances exist at lower levels. The zodiacal light and transient plasma clouds in the solar wind both make contributions (Richter *et al.*, 1982), and scattered light both from the Sun and from nearby planetary objects represents a design problem.

8. CONCLUSIONS

This paper has discussed ideas for a sky survey devoted to the measurement of photometric variability. This kind of observation would enable us to study low-level variability beyond the capability of Earth-based observations, and to do it for faint as well as bright objects. We would obtain extended time series, permitting the measurement of non-radial oscillations main-sequence and other stars across some large fraction of the sky. The data base generated would provide photometric information for the entire sky, down to a limiting magnitude of 12 in a three years' mission. It would provide deeper surveys in smaller fractions of the sky, to about 18th magnitude in about 10% of the sky per year.

As mentioned in the introduction, there are gross technical conflicts obstructing goal of obtaining broad, continuous, precise coverage of many astronomical objects with different brightnesses scattered across the whole sky. This means that compromises are necessary. Hudson (1984) gives one version of these tradeoffs based upon the APE concept described above more briefly. It would be very desirable to continue to sharpen the choices and to explore alternative concepts, such as the "planetarium camera" mentioned above, or other ideas presented in the Meudon workshop (Mangenet and Praderie, 1984).

The photometry of bright objects is presently limited by systematic effects on the Earth, and will be limited by (other) systematic effects in space. The nature of these new problems depends upon the kind of photometry one wants to do - many of the effects, such as detector and filter drift, are already recognized. The results from the Hubble Space Telescope will be interesting in that we will have a much better idea about the nature of photometric limitations to several kinds of detector. We should make sure that this kind of experience and that gleaned from other, simpler, space experiments, finds its way into planning for any future photometric variability survey.

Acknowledgments: The Steward Observatory group suggested the idea of a "planetarium camera." Most of this paper originally appeared in the workshop proceedings "Space Research Prospects in Stellar Activity and Variability," edited by A. Mangenet and F. Praderie.

REFERENCES

Christensen-Dalsgaard, J., and Frandsen, S., 1983, **Solar Phys.**, 82, 469.

Foukal, P.V., and Vernazza, J., 1979, **Astrophys. J.**, 234, 707.

Gough, D., 1984, **Advances in Space Research**, to be published.

Hudson, H., 1984, in A. Mangenet and F. Praderie (Eds.), **Space Research Prospects in Stellar Activity and Variability**, (Observatoire de Paris, Meudon), p. 197.

King, I.R., 1983, **Pub. Ast. Soc. Pacific**, 95, 163.

Kurtz, D.W., 1982, **Monthly Notices of the Royal Astronomical Society**, 200, 807.

Livingston, W.C., and Ye, B., 1982, **Publ. Ast. Soc. Pacific**, 94, 713.

Mackay, C.D., 1982, **Proc. S.P.I.E.**, 331, 146.

Mangenet, A., and Praderie, F., 1984, in A. Mangenet and F. Praderie (Eds.), **Space Research Prospects in Stellar Activity and Variability**, (Observatoire de Paris, Meudon), p. 379.

McGraw, J.T., Stockman, H.S., Angel, J.R.P., and Epps, H., and Williams, J.T., 1982, **S.P.I.E. 331**, 137.

Richter, I., Leinert, C., and Planc, B., 1982, **Astron. Astrophys.**, 110, 115.

Ricker, G.B., Doty, J.P., Vallerga, J.V., and Vanderspek, R.K., 1984, in S. Woosley (Ed.), **High Energy Transients in Astrophysics** (New York, American Institute of Physics), p. 669.

Sterne, T., and Dieter, N., 1958, **Smithsonian Contributions in Astrophysics**, 3, No. 3, 9.

Walker, A., 1984, **Mon. Not. Roy. Astr. Soc.**, 209, 83.

Westphal, J.A., 1982, in D.N.B. Hall (ed.), **The Space Telescope Observatory** (NASA CP-2244), p. 28.

Willson, R.C., 1984, in B. LaBonte, G. Chapman, H. Hudson, and R. Willson (Eds.), **Solar Variability on Active-Region Time Scales** (NASA CP-2310), p. 1.

Willson, R.C., Gulkis, S., Janssen, M., Hudson, H.S., and Chapman, G.A., 1981, **Science**, 211, 700.

Woodard, M.F., and Hudson, H.S., 1983, **Nature**, **305**, 589.

Woodard, M.F., Hudson, H.S., and Willson, R.C., 1982, in J.P. Cox and C.J. Hansen (Eds.), **Pulsation in Classical and Cataclysmic Variable Stars** (Boulder, JILA), p. 152.

Young, A.T., 1974, in N. Carleton, (ed.), **Methods of Experimental Physics**, vol. 12.

II. ATMOSPHERIC EFFECTS

THE GROUND BASED PHOTOMETRIC LIMITATIONS TO THE SEARCH FOR LIGHT VARIATIONS DUE TO 5-MINUTE SOLAR-TYPE OSCILLATIONS IN OTHER STARS

D. W. Kurtz
Department of Astronomy, University of Cape Town

Summary. High speed photometric observations obtained with a single channel photometer on the SAAO 0.5-m and 1.9-m telescopes are presented. It is argued that the dominant source of noise at periods near 5 minutes is sky transparency variation rather than scintillation. For bright stars this means that, at this period, increased telescope aperture does not improve photometric accuracy. It is claimed that ground based photometric observations cannot reach accuracies sufficient to study solar-type oscillations in other stars. Such observations must be made from space. Recommendations for improving ground based observations are made.

It is now well accepted that the sun oscillates in many high order, low degree p-modes with periods near 5 minutes. The study of the frequency spectrum of these oscillations gives observational constraints to the theory of the solar interior. Enthusiasm for solar seismology has led many people to wish to search for similar oscillations in other stars through the photometric monitoring of their brightnesses. Woodard and Hudson (1983) have shown that the individual oscillation modes in the sun have amplitudes in the range of $2\text{-}4\times10^{-6}$ of the mean irradiance. Theoretical predictions of the amplitudes expected for similar oscillations in other stars have been as much as 2×10^{-5} of the mean brightness (Christensen-Dalsgaard & Frandsen 1983). Deubner & Isserstedt (1983) have made photometric observations of two solar-type stars and claimed that the upper limit to any oscillations with periods near 5 minutes that may be present in either of these two stars is about 5.5×10^{-5}; they are, therefore, optimistic about the chances of detecting such oscillations with ground based photometry.

For the last five years I have been making intensive photometric observations of light variability in the period range 5 to 15 minutes in cool Ap stars (Kurtz 1982). These observations have been obtained primarily at the South African Astronomical Observatory (SAAO), an excellent photometric observing site, and they have been obtained with well tested, stable, reliable, conventional single-channel astronomical photometers. In the period range of 5 to 15 minutes they are among the most accurate photometric observations ever made.

Fig. 1 shows a 5.4-hr long light curve of HD 101065 obtained on 1984 March 26/27 through a Johnson B filter using the SAAO 0.5-m telescope and Peoples Photometer with 20-s integrations. The observations have been corrected for mean extinction and have had some long period (P > 2 hr) noise filtered out. The 12.14-minute

oscillations can be easily seen as can the $5x10^{-3}$ (5 mmag) high frequency scatter. In my experience these observations are typical of the best that can be obtained with an 0.5-m telescope with a good photometer and from a good site. An amplitude spectrum of the light curve shown in fig. 1 can be seen in fig. 2. The peak at a period of 12.14 minutes (f=4.94/hr) dominates the amplitude spectrum. The noise level from 3/hr out to the Nyquist frequency of 90/hr is about $4x10^{-4}$ (0.4 mmag) and is traditionally presumed to be due to scintillation. The low frequency noise is substantially greater than this (even though some very low frequency noise has been filtered out already) and is traditionally attributed to sky transparency variations. Without any simultaneous comparison star observations, the source of this low frequency noise cannot be determined and hence it is usually ignored. The count rate for the observations presented here is 800,000 per 20-s integration which means that the photon noise is substantially less than the scintillation noise.

With 5 hours of observation with an 0.5-m telescope on a bright star I find the noise in the amplitude spectrum at periods near 5 minutes to be $4x10^{-4}$ (0.4 mmag). By putting together 138 hours of observation on 38 nights (see the discussion of HR 3831, Kurtz 1982), I have reduced the noise at a period of 5 minutes to $1x10^{-4}$ (0.1 mmag). This is two orders of magnitude too great to study solar-type oscillations which have amplitudes in the range of $2\text{-}4x10^{-6}$. Deubner & Isserstedt (1983) claimed that they had a detection limit of $5.5x10^{-5}$ using techniques similar to those presented here. They were not trying to study individual modes, however, they were trying to detect the presence of 80 modes equally spaced in frequency as is seen in the sun. They therefore averaged their amplitude spectra 40 times modulated by trial frequency separations. For comparison their noise level should be multiplied by $\sqrt{40}$ which gives $3.5x10^{-4}$. This is comparable with the noise shown in fig. 2 although Deubner & Isserstedt used about 35 hr of observations.

The flat noise spectrum shown in fig. 2 is traditionally ascribed to scintillation noise even though studies of scintillation have generally been confined to frequencies greater than 1 Hz. Scintillation noise is well known to decrease roughly proportionally to increasing telescope aperture so, in principle, it should be possible to reduce the noise level in fig. 2 by a factor of 10 to $4x10^{-5}$ (0.04 mmag) with an equivalent number of observations with a 5-m telescope.

In order to test this, I obtained photometric observations of HD 101065 with the SAAO 1.9-m telescope. Fig. 3 shows a 9.1-hr light curve obtained on 1984 March 14/15. By comparison with fig. 1 it can be seen that the scintillation noise has been dramatically reduced due to the greater aperture of the 1.9-m telescope. The count rate has also increased, of course, to 10^7 per 20-s integration, but that is not the major cause of the difference in the two diagrams. I judged the photometric quality of the two nights to be similar.

Fig. 4 shows an amplitude spectrum of the light curve given in fig. 3. As was expected, the scintillation noise at the high frequency end of the spectrum is substantially lower for data obtained with the 1.9-m telescope than for data obtained with the 0.5-m telescope. If we consider the difference in the duration of the observations in figs. 1 and 3, then the relative amplitude of the high

frequency noise in figs. 2 and 4 is, within the errors, equal to the ratio of the telescope apertures, a predictable result. However, a closer scrutiny of figs. 2 and 4 near a period of 5 minutes (f=12/hr) indicates that the noise level at that period is comparable for the two telescopes. I conclude from this that photometric noise near periods of 5 minutes is not primarily scintillation noise since it is relatively insensitive to aperture.

This point deserves further examination. Fig. 5 is an amplitude spectrum of two consecutive nights of observation of HD 101065 with the 0.5-m SAAO telescope and fig. 6 is an amplitude spectrum of two similar nights of observation with the 1.9-m telescope. The 0.5-m telescope observations have a duration of 16.3 hr, the 1.9-m telescope observations, 15.4 hr. All four nights were of good photometric quality. It can be seen that at a period of 2 minutes (f=30/hr) the 1.9-m telescope data are clearly superior to the 0.5-m telescope data, whereas at a period of 5 minutes the 1.9-m telescope data are only marginally better. At longer periods the larger aperture confers no advantage.

(Parenthetically I must announce a caveat. The photometer on the SAAO 0.5-m telescope is a dedicated instrument; it is always on the telescope and always in use. It is, as a consequence, continually tested and maintained and is highly dependable. The photometer which I used on the 1.9-m telescope is only occasionally in use. During the time when the observations presented in this paper were gathered, there were some small, but noticeable, sensitivity variations across the 30 arcsec observing aperture. These variations were minimized through careful tracking and guiding, but undoubtably contribute some noise at very low frequency to the amplitude spectra presented in figs. 4 and 6. I judge that such sensitivity drifts are not, however, a major contributor to the noise at periods near 5 minutes and hence do not affect the conclusion that such noise is relatively insensitive to telescope aperture.)

In fig. 7 I present a speculative diagram of the sources of photometric noise at frequencies of interest in the search for solar-type oscillations in other stars. At periods of a few minutes and less the noise is probably due to scintillation and appears to be flat. At periods longer than a few minutes the noise rises towards longer periods and is probably due to sky transparency and sky backgroud variations. Many astronomical photometrists are are well aware of the dependency of the observed count rate on humidity and dust levels which can vary during the night. From day time observations (Clarke 1978) there is evidence for regular or semi-regular variations in sky brightness on time scales of tens of minutes. Such variations should also occur on moonlit nights. Short time scale barometric oscillations are known which may affect astronomical transparency. Night sky emission lines may also vary substantially on a time scale of minutes.

Thus a major conclusion of this paper is that in the use of high speed photometric techniques with single channel photometers to study the light variations in ZZ Ceti stars, Rapidly Oscillating Ap stars, or any other astronomical object which varies or may vary on a similar timescale, large aperture telescopes have little advantage over small aperture telescopes for bright objects. The largest telescope needed is one which reduces photon counting noise (for integration times half

the length of the period of interest) below sky transparency noise.

In principle, the use of two- or three-channel photometers should make it possible to eliminate sky transparency and sky background variations to produce scintillation limited amplitude spectra at all frequencies. (It should be noted that we do not know anything about scintillation noise at periods of several minutes and longer.) To my knowledge, no one has ever done this for high speed photometry. Evidence for this comes largely from the lack of any observations of higher accuracy than the observations of the Rapidly Oscillating Ap stars which were obtained with single channel photometers. At the very lowest frequencies, of course, all photometrists reduce sky transparency variations through the use of comparison stars.

In practice two-channel photometers have been used to improve poor quality data obtained under marginal observing conditions rather than to improve high quality data. One of the reasons for this is that for bright stars with low photon noise, it is usually difficult to acquire a comparison star of similar brightness in the limited field of the second channel. Hence, if the program star data are normalized to the comparison star data on a point-by-point basis, the larger photon noise of the comparison star data degrades the quality of the program star data. Observers using two-channel photometers have in general examined comparison star data from the second channel and then discarded them on good photometric nights. I suggest that users of two-channel systems should smooth their comparison star data over half the period of interest in the program star and then normalize the program star data with the comparison star data (assuming that the photon noise for the comparison star is less than sky transparency noise when integrated over half the period of interest in the program star). Normalization can also be done in frequency space with the real and imaginary parts of the Fourier transforms of the program and comparison star light curves.

For the highest accuracy observations it may be necessary to have a third channel to monitor sky background variations. Clarke (1978) claims 1% sky brightness variations during the day which should also be present on moonlit nights. As an example, typical observing conditions with the moon up might result in a sky background count rate which is 10% of the count rate on the star. A 1% variation in that sky background will produce 1×10^{-3} (1 mmag) noise in the stellar data. With differences in brightness of program and comparison stars, one can easily see this as a source of noise substantially above the 1×10^{-4} (0.1 mmag) limit obtained already at higher frequencies.

It seems possible that a three channel photometer may be able to reduce noise at a period of 5 minutes to the level of scintillation noise. It is also possible that the use of a CCD detector may allow this (Walker, this conference). Both of these conjectures remain to be proved. Data shown in this paper, however, lead to the conclusion that scintillation limited observations will only reach a noise level of at best 3×10^{-5} with one night of observing with a 5-m telescope. With 100 closely spaced nights of observing with a 5-m telescope (clearly unavailable) this limit might be reduced to the amplitude of the solar irradiance variations. That would only allow the detection of such variations in other stars but not the study of them because of the poor 1-to-1 signal-to-noise ratio. It may be possible to detect solar-type oscillations in other stars by averaging amplitude spectra many times as Deubner & Isserstedt tried, but the study of such

oscillations cannot be done by ground-based observation. Such a study will only be possible with a dedicated orbiting telescope.

I would like to make some miscellaneous points here which are related to the above discussion. 1) Observations of Rapidly Oscillating Ap stars show that the phase of the oscillations in those stars is strongly dependent on wavelength in some cases. In HR 3831 the phase of the principal frequency is shifted 26 degrees between U and B and a further 28 degrees between B and V (Kurtz 1982). Theoretical considerations (Balona 1981) lead to the expectation of these phases shifts for high overtone p-modes, although quantitative agreement between observation and theory is not yet good. If white light observations of these stars were made, phase smearing would drastically reduce the signal-to-noise ratio. It would be interesting to know what the solar irradiance variations are like when observed through narrow band filters. Astronomers planning to search for solar-type oscillations in other stars either from the ground or space should consider whether white or filtered light observations will provide the highest signal-to-noise ratio.

2) Astronomical phototubes are generally limited to count rates of about 10^6/s both because of dead-time losses and due to tube damage at higher rates. This is not a limitation for ground based photometry since at a count rate of 10^6/s photon noise is less than scintillation noise. However, count rates of 10^7/s to 10^8/s may be available to an orbiting telescope where scintillation noise is non-existent. Phototubes which can take advantage of such high count rates would be desirable in order to obtain observations of the highest possible quality.

REFERENCES

Balona, L. A., 1981. Mon. Not. R. astr. Soc., 196, 159.

Christensen-Dalsgaard, J., & Frandsen, S., 1983. in IAU coll. 66, p. 469, ed. D. O. Gough, (Reidel).

Clarke, D., 1978. Phys. Bull., 29, 505.

Deubner, F.-L., & Isserstedt, J., 1983. Astr. Astrophys., 126, 216.

Kurtz, D. W., 1982. Mon. Not. R. astr. Soc., 200, 807.

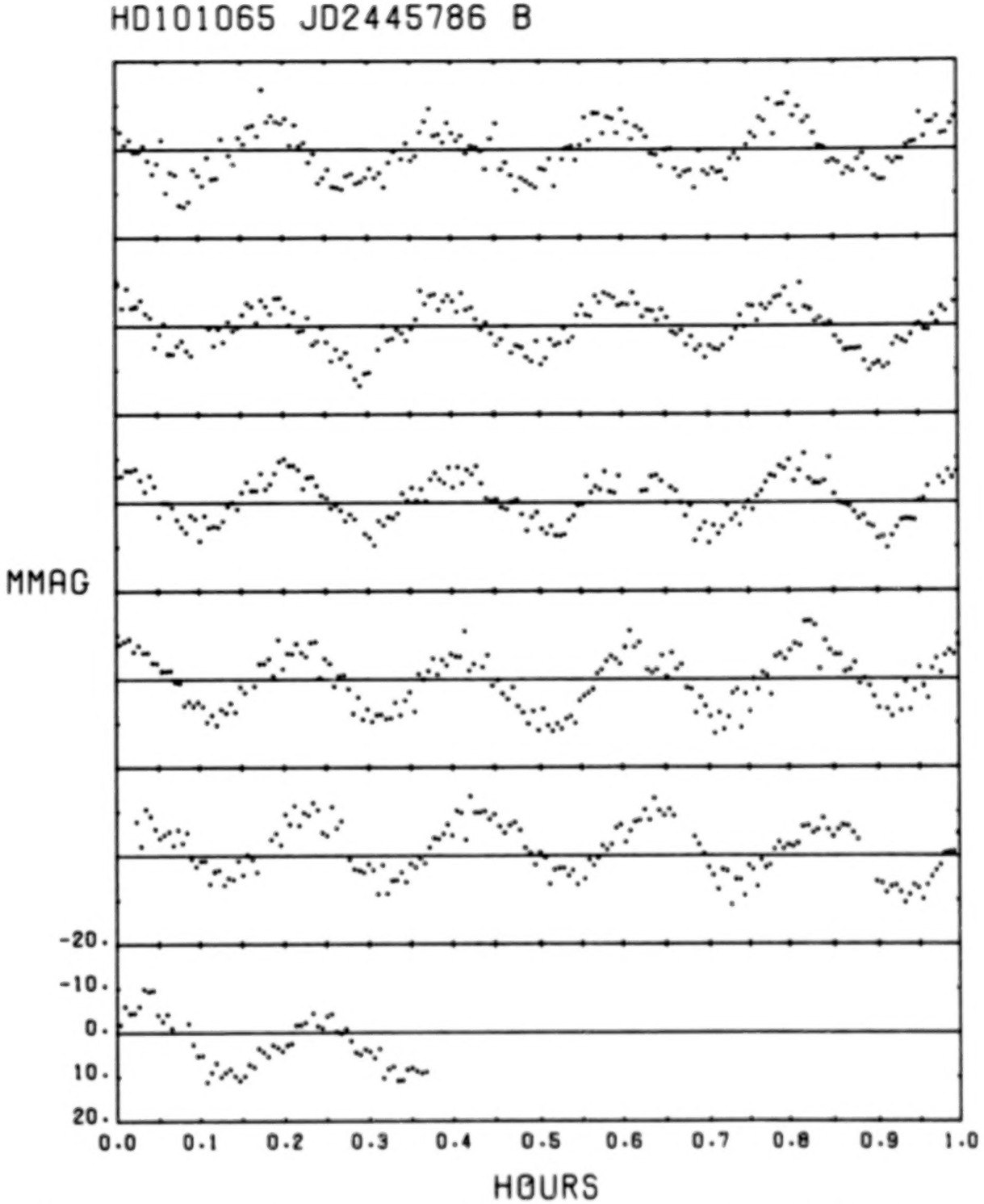


Figure. 1. The B light curve of HD 101065 observed on JD2445786 with the SAAO 0.5-m telescope. This plot as well as fig. 3 has the following construction. The abscissa is 1 hr long and the ordinate is 40 mmag high for each panel with tick marks at 10 mmag intervals. The continuous 5.37-hr long light curve has been folded into 1-hr long sections which should be read left-to-right, top-to-bottom. The values of the ordinate are the same in all panels although they are labelled only for the lowest panel.

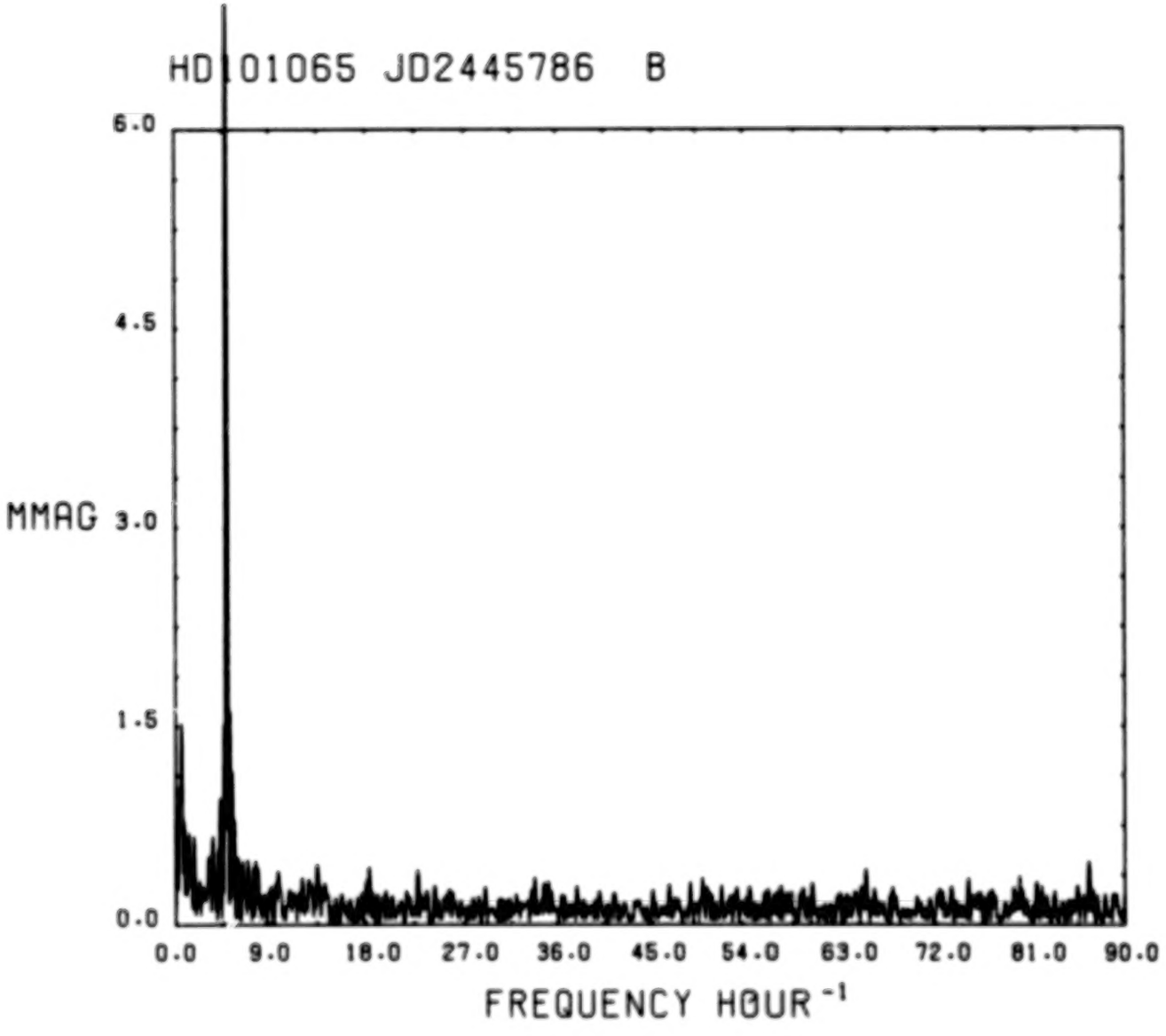


Figure 2. The amplitude spectrum of the light curve shown in fig. 1. Here amplitude is semi-amplitude in the relation $A\cos(\omega t+\phi)$.

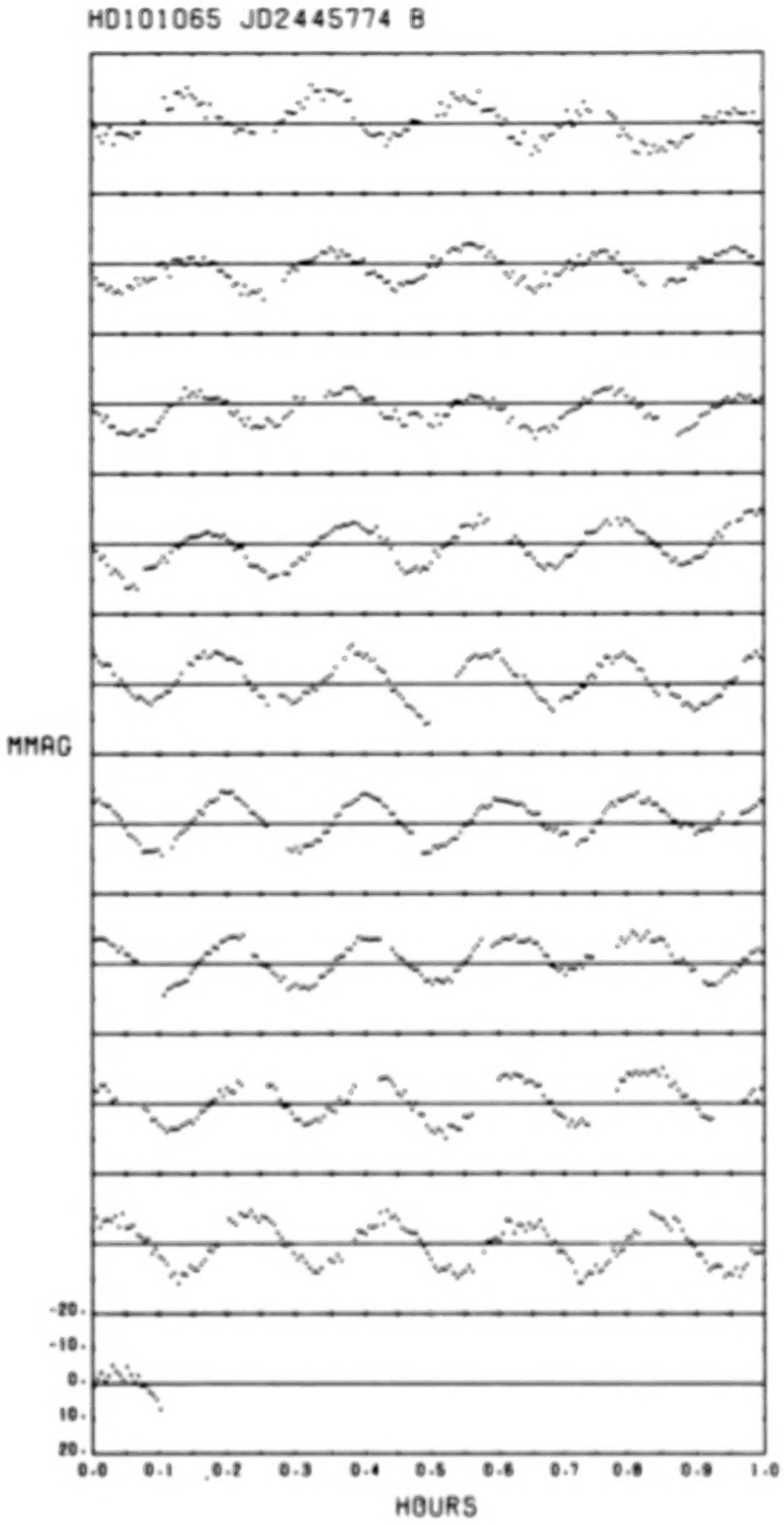


Figure 3. The B light curve of HD 101065 observed on JD2445774 with the SAAO 1.9-m telescope. This is a continuous 9.1-hr long light curve with the same construction as the light curve in fig. 1.

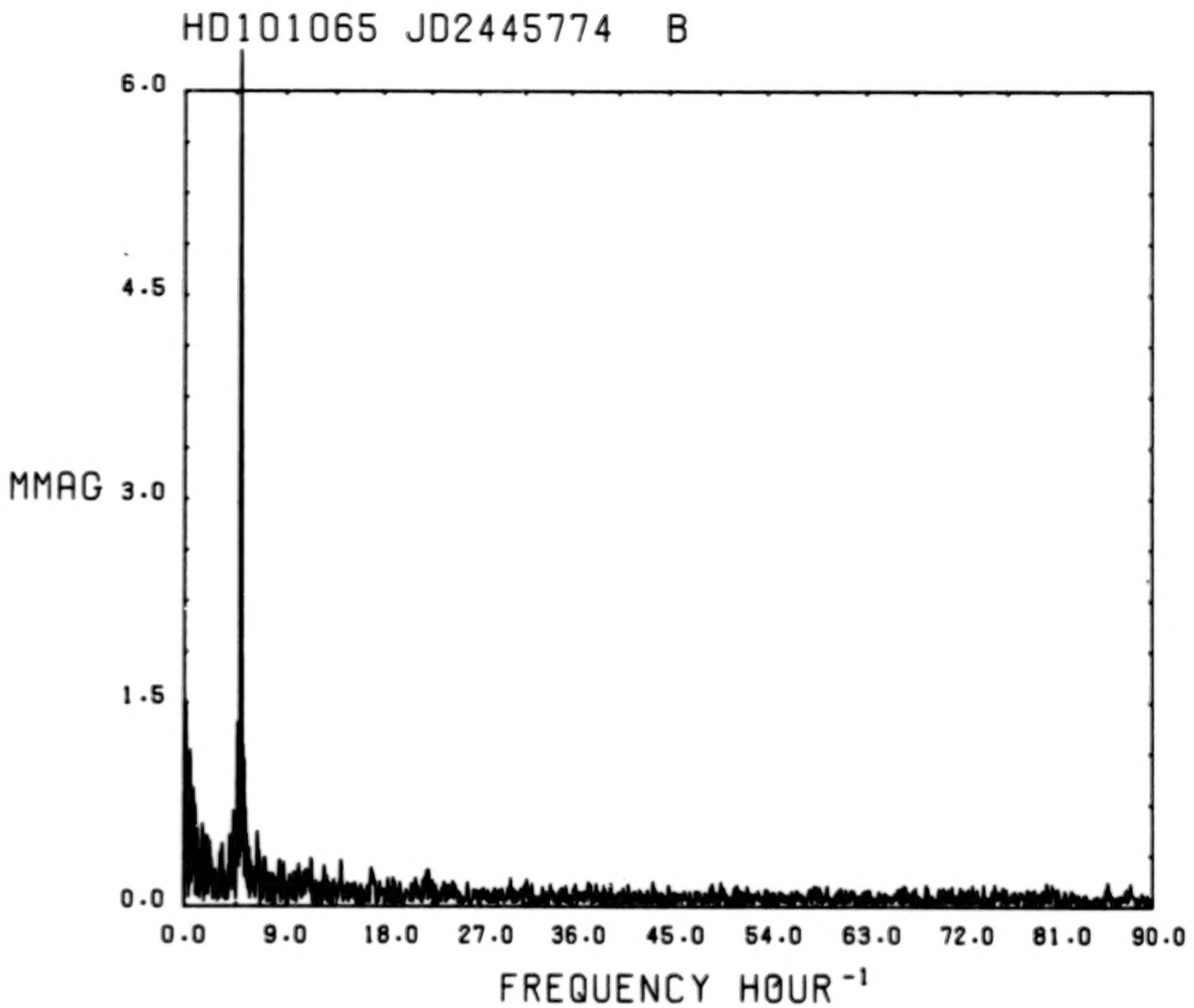


Figure 4. The amplitude spectrum of the light curve shown in fig. 3.

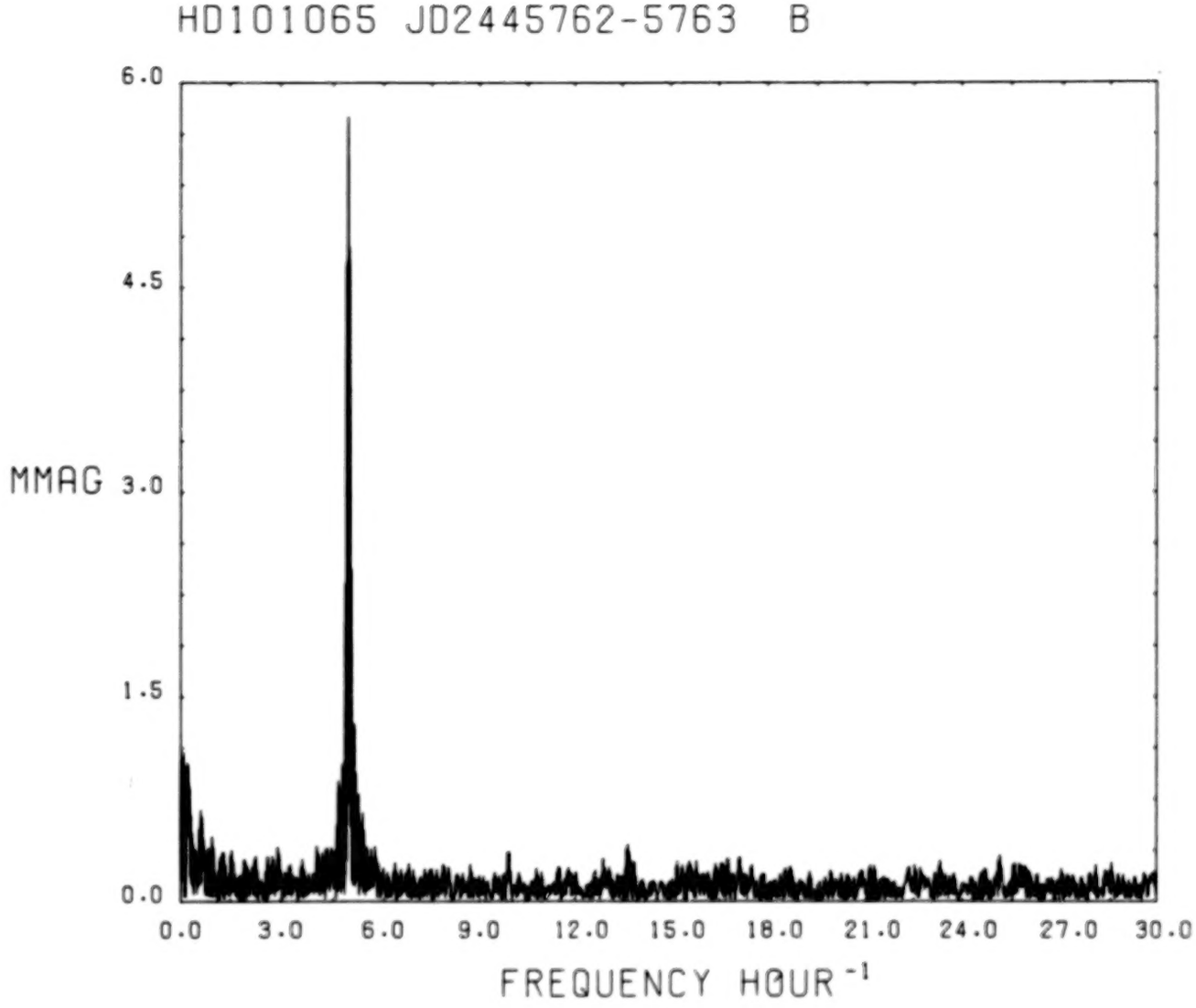


Figure 5. The amplitude spectrum of 16.3 hr of observations of HD 101065 obtained with the SAAO 0.5-m telescope on two consecutive nights.

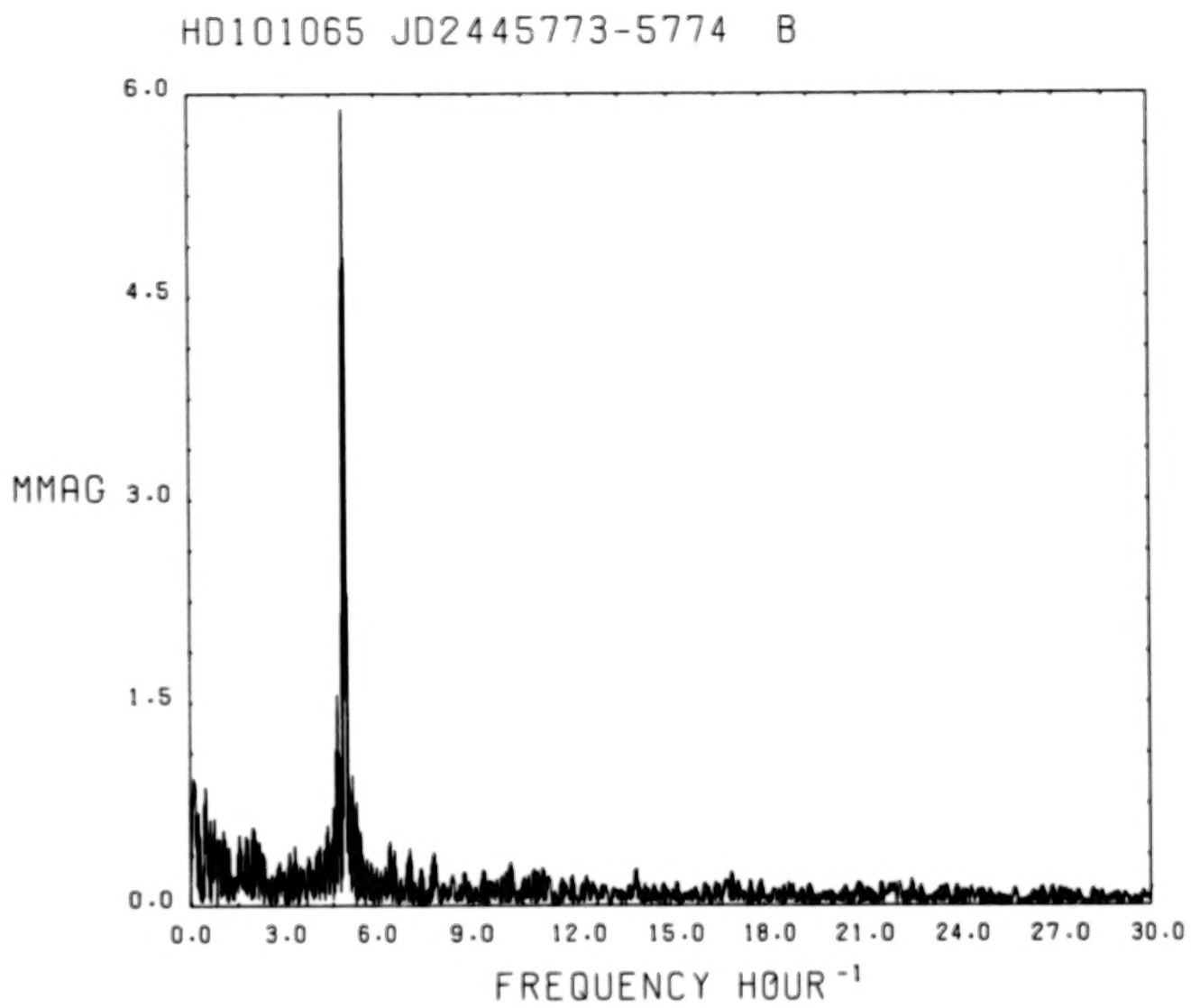


Figure 6. The amplitude spectrum of 15.4 hr of observations of HD 101065 obtained with the SAAO 1.9-m telescope on two consecutive nights.

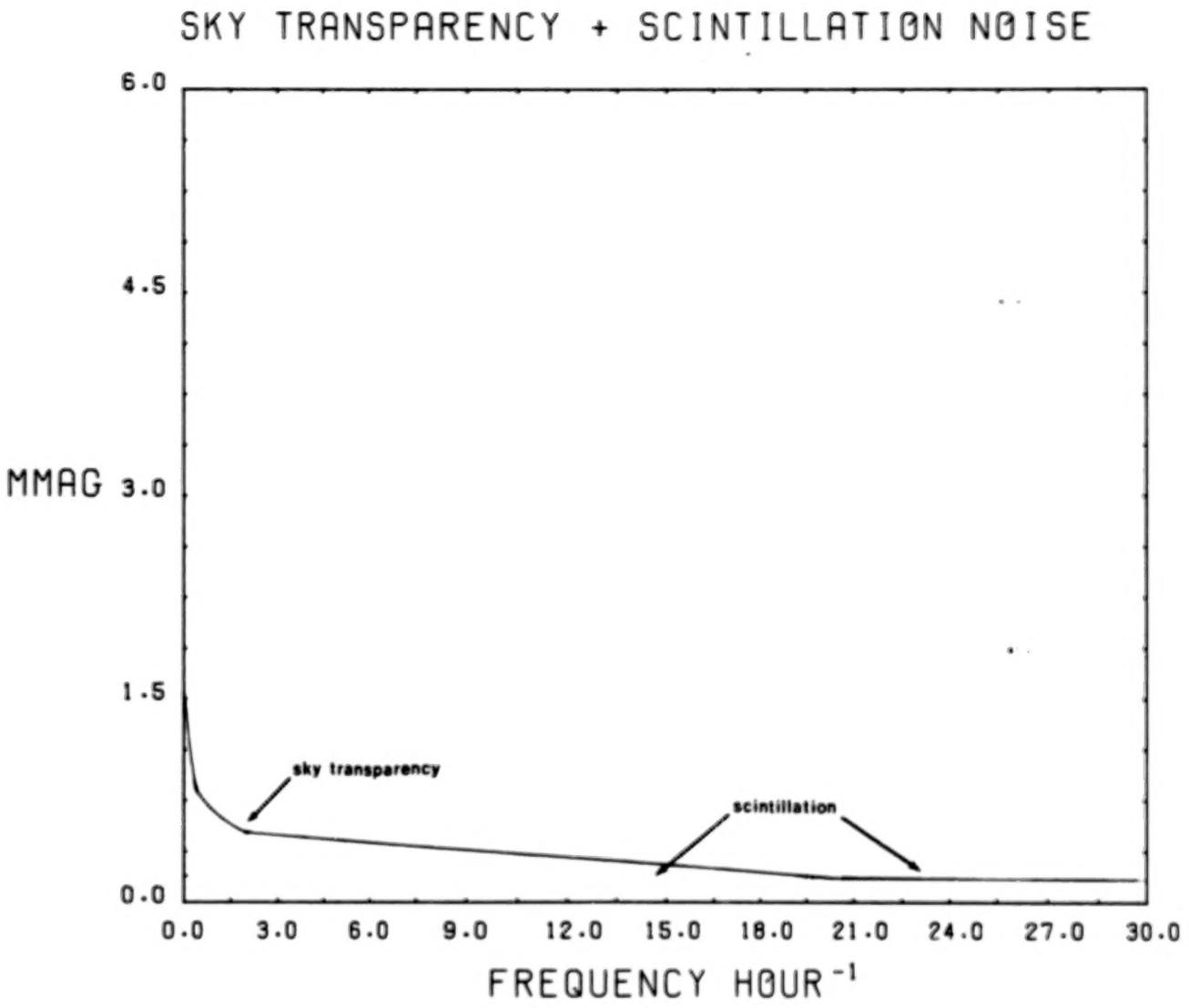


Figure 7. A schematic amplitude spectrum showing the suggested shapes of the noise distributions due to sky transparency variations and scintillation. In practice the low frequency noise also contains contributions from sky background variations and from instrumental variations. It is not known whether two- or three-channel photometers can be used to reduce the low frequency noise to the level of the scintillation noise. It is also not known what the level of the scintillation noise is at low frequencies.

D7

ATMOSPHERIC LIMITAIONS IN STELLAR SEISMOLOGY
SHOULD ONE MEASURE RADIAL VELOCITY OR BRIGHTNESS FLUCTUATIONS?

Eric FOSSAT
Observatoire de Nice and
Departement d'Astrophysique de
l'Universite de Nice
Parc Valrose
F - 06034 NICE CEDEX

1 - Introduction. Low degree p-modes of the Sun have been measured in spatially integrated sunlight (the Sun as a star) both in Doppler shift and in intensity fluctuations. (An extended bibliography can be found in the proceedings of several recent meetings, held in Crimea in September, 1981, at Catania in May, 1983 and at Snowmass in August, 1983). These observations are a good starting point for the discussion of the best way to collect equivalent data on other stars. The guideline of this paper will be to assume that the Sun is removed far enough is space to become an ordinary star of magnitude zero to one. Evidently, another star will oscillate with different frequencies and different amplitudes, but one must use some reference to start with. Using this scheme, a detailed investigation of the limitations of observational accuracy in the search for global p-modes is made. The sources of noise stand in the Sun itself, in the instrumentation, in the observing time duration, in the corpuscular nature of the light and mostly in the Earth atmosphere in the case of ground based observations.

2 - Need for long time series. The effect of observing time duration is illustrated by the first four figures. Both resolution and sensitivity are improved with the increase of the time series length and this example shows that in the solar case, a one-day (or one night for an equivalent star) integration time is long enough to resolve the discrete structure of the power spectrum and to measure the mean frequency spacing $\Delta\nu$. The identification of every single mode requires more resolution, obtained typically with a few days of data. It has been shown (J. Connes 1984) that to a certain extent, this data needs not to be continuous, provided the coherence time of each oscillation is much longer than the data gaps.

For a stellar observation, it can be deduced that the frequency resolution provided by one typical photometric night is presumably adequate for p-mode detection. Then, a good set of several continuous nights should be long enough (with some mathematical work for filling the gaps) to obtain a resolution comparable with the one in Figure 3, which makes possible a complete eigenmode identification. The next steps, which are the study of amplitude time variations and the measurement of rotational

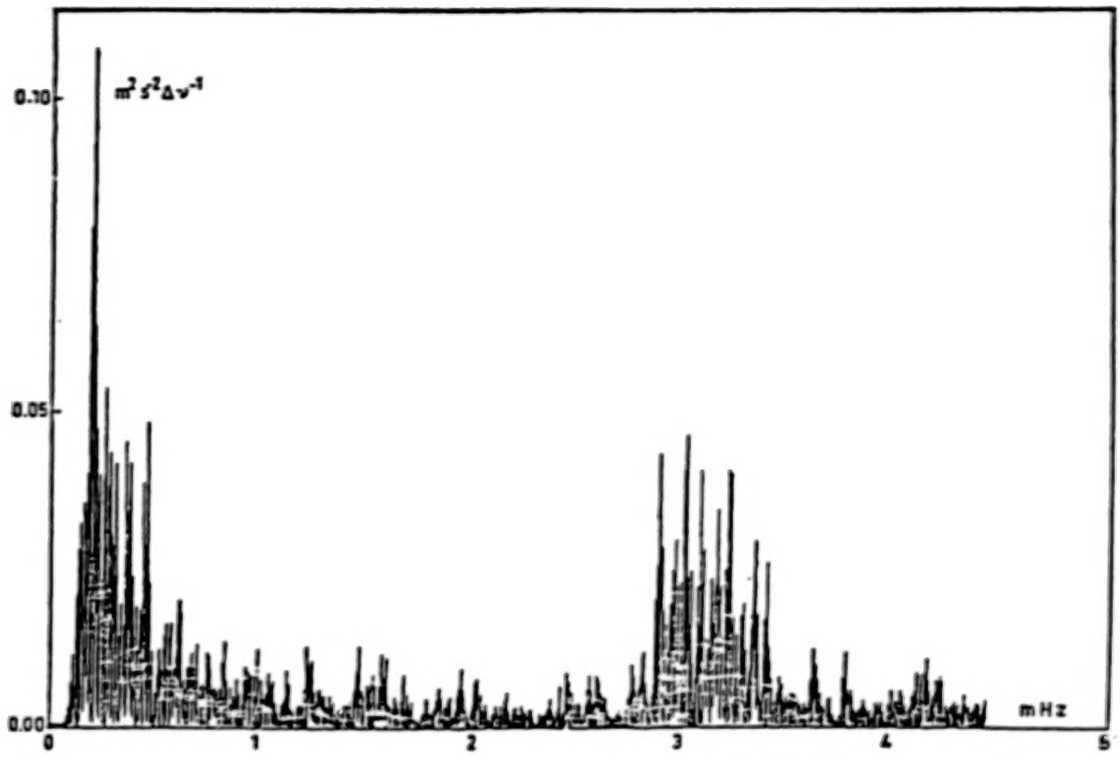


Figure 1- (from Grec and Fossat, 1976). Full disk solar Doppler data obtained during 4 out of five consecutive days at Nice in October 1975 provided this power spectrum. The discrete spiked structure is visible in the five minute range (3 mHz) but was not regarded as significant.

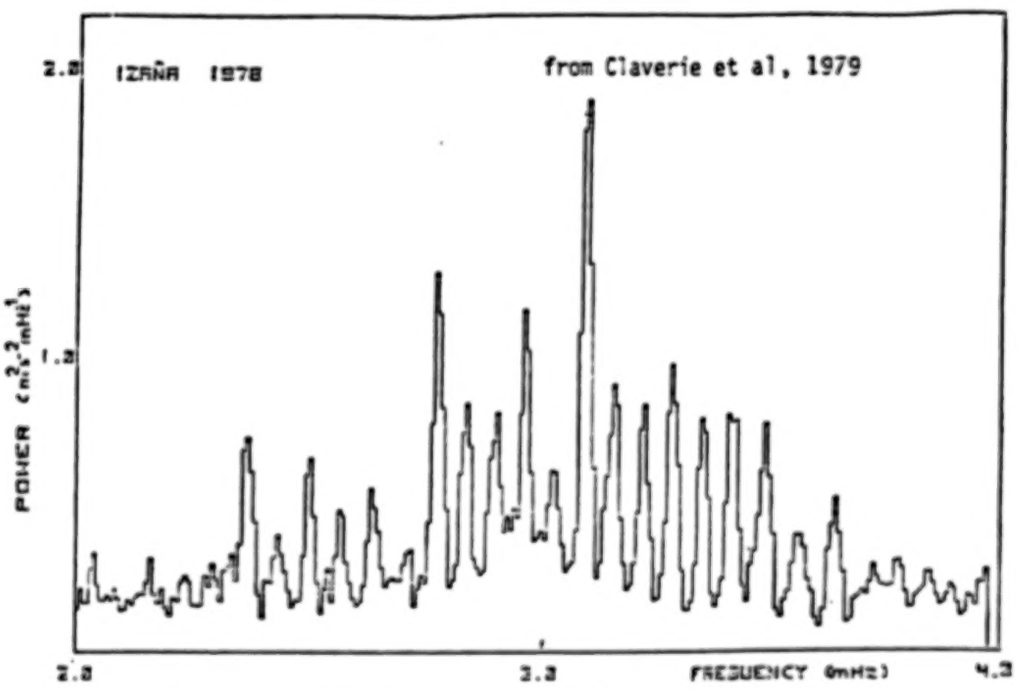


Figure 2- First published power spectrum interpreted as evidence for global low degree solar p-modes.

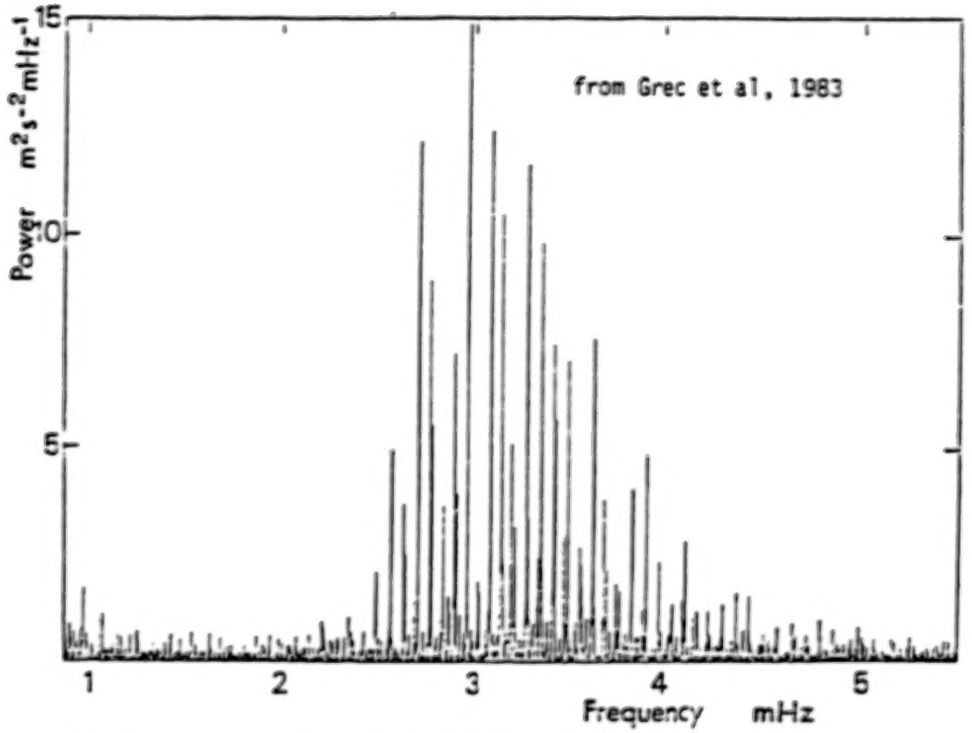


Figure 3- Same as figures 1 and 2, from 6 days of data obtained at the Geographic South Pole in January 1980

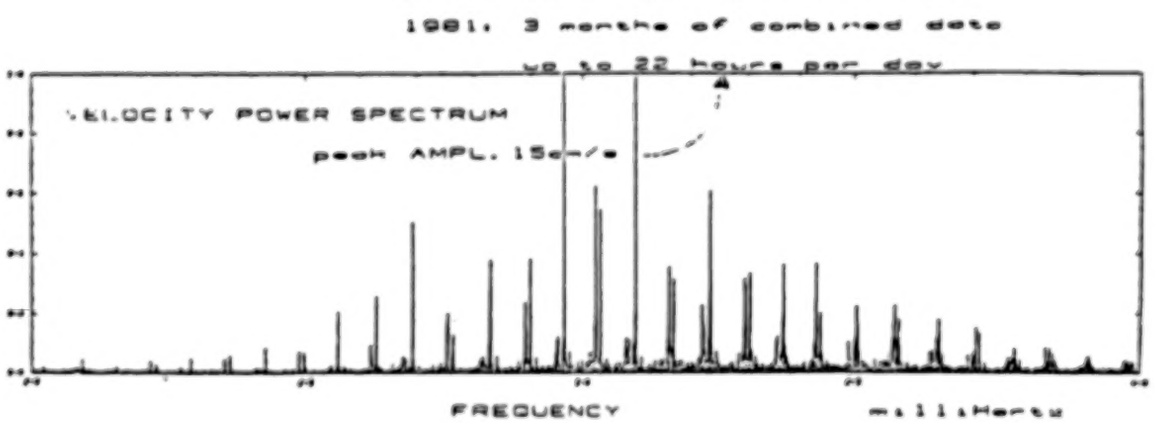


Figure 4- (from the Birmingham group). Same as figures 1, 2 and 3 from almost three months of combined data obtained at Teneriffe and at Hawaii.

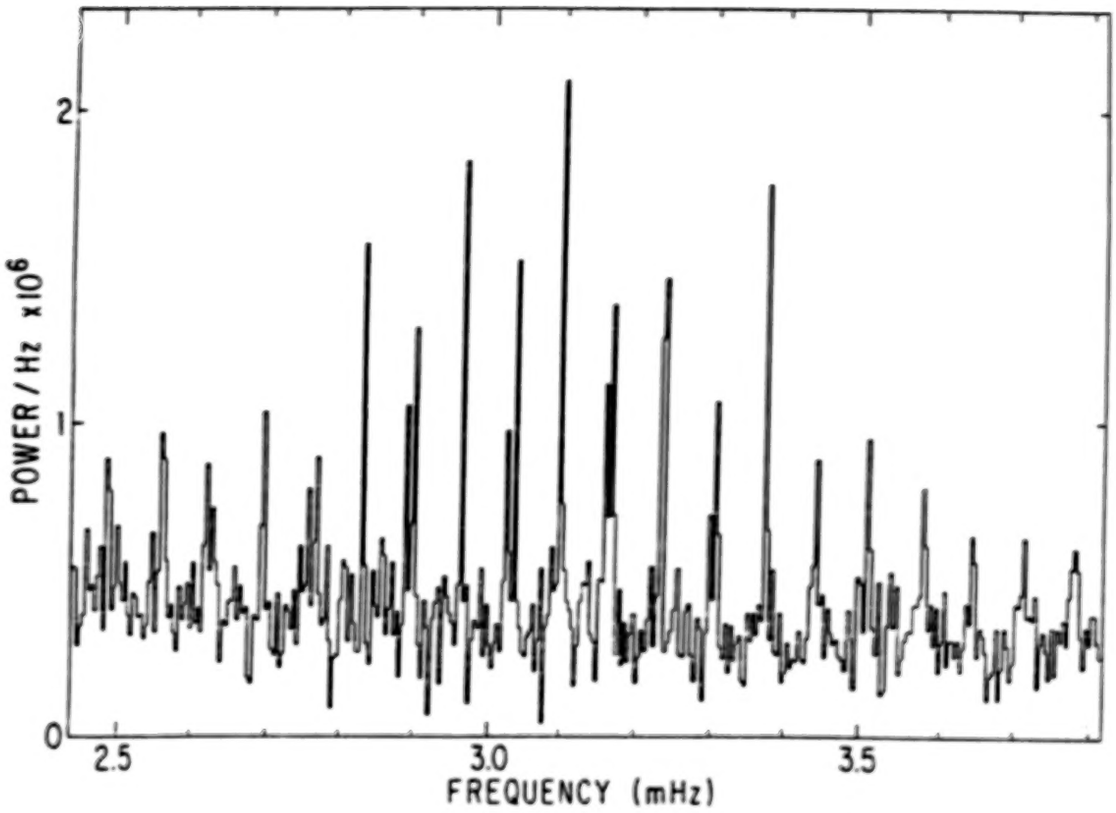


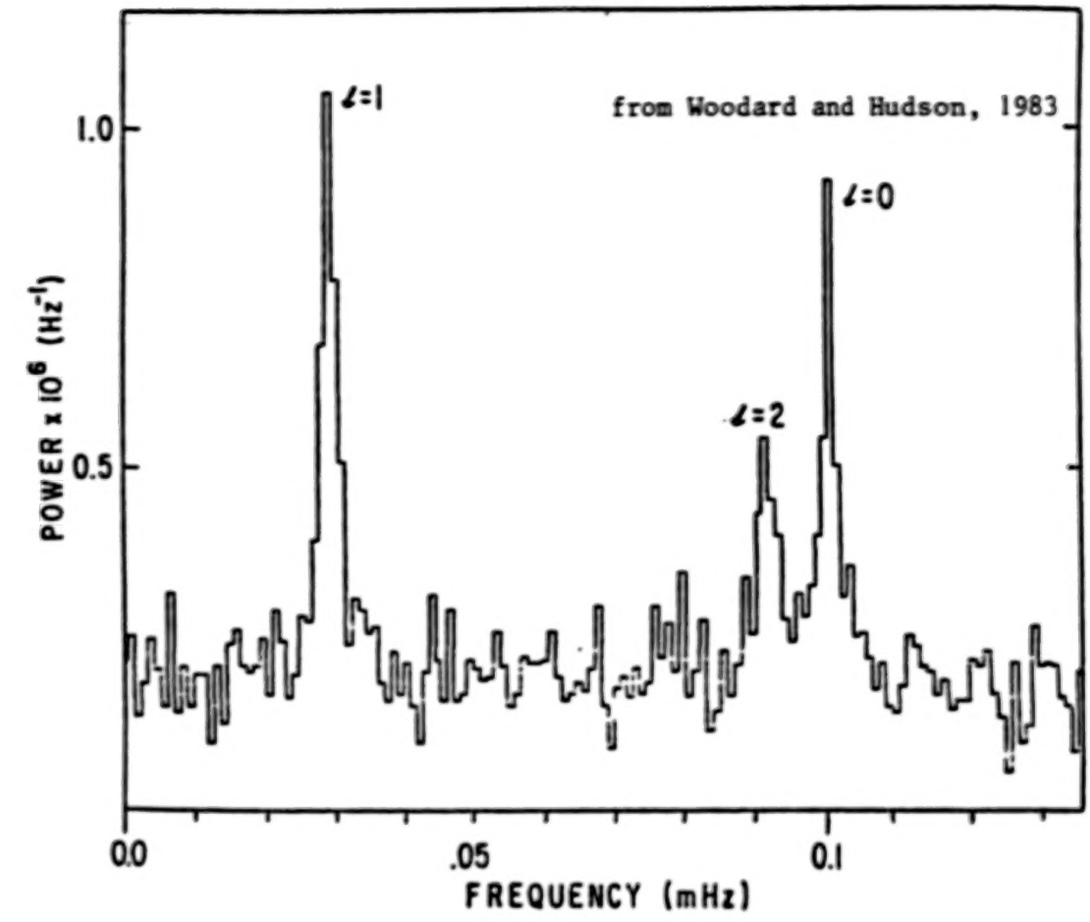
Figure 5: (From Woodard and Hudson, 1982). Same as Figures 1 to 4 from eight months of global irradiance data obtained with the ACRIM experiment onboard the SMM spacecraft.

splittings, will require much longer and more continuous data sets, clearly unavailable in one single site, except for the Geographic South Pole.

3 - Stellar Noise. For comparison with these first four figures obtained from full disk Doppler shift measurements, Figure 5 shows the same p-mode spectrum obtained from bolometric brightness measurements made by the ACRIM instrument onboard the SMM spacecraft.

With an adequate folding and averaging (superposed frequency analysis), the mean shape and the mean power of modes of each angular degree can be displayed. This is shown in Figure 6 for Doppler shift and brightness measurements. Note the power scales that give the mean amplitude of individual modes and which will be used somewhat further for comparison with atmospheric noise.

Note also that the mean ratio between the peak heights and the background level is about 20 in Doppler shift , but only 3 to 4 in brightness. This background level is due to instrumental noise, atmospheric noise (for Doppler shift measurements only), photon statistics, higher degree unresolved eigenmodes and convective solar noise. In my opinion, this last contribution is very likely to be more important in brightness measurements than it is in Doppler shifts. Indeed, the r.m.s.



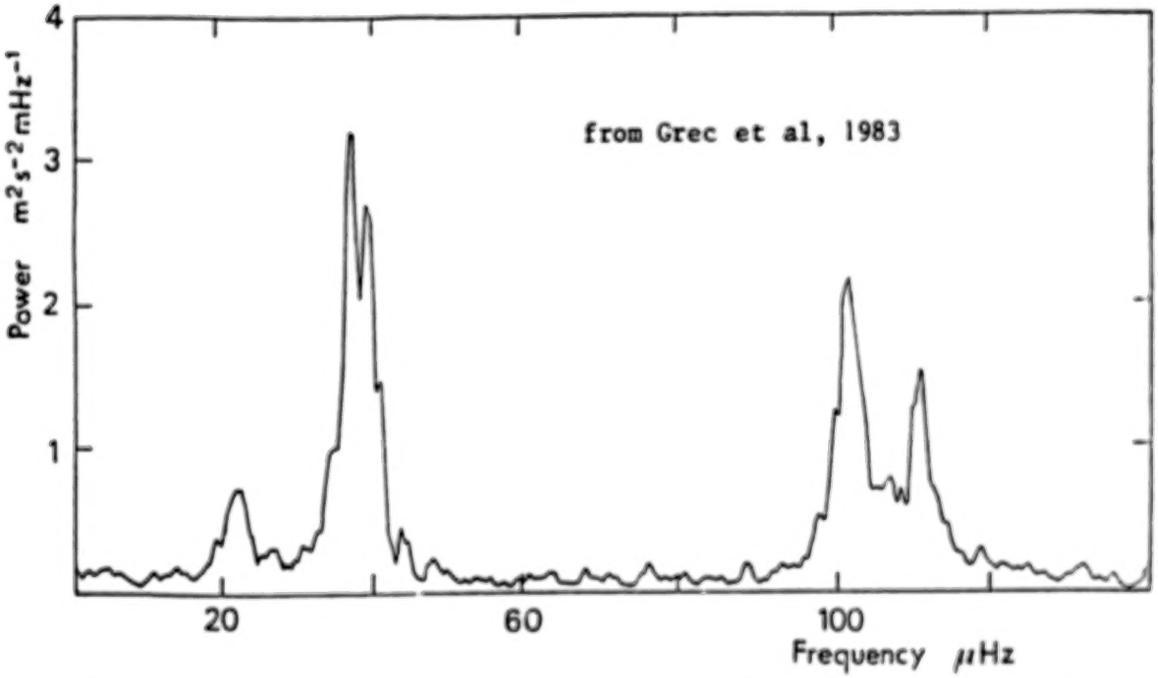


Figure 6 - A superposed frequency analysis of a power spectrum like those of the previous figures displays the mean shape of individual peaks. This being done for Doppler shift or brightness data shows the difference of relative sensitivity of these two types of observation to the degree of the eigenmodes.

contrast of white light solar granulation is of the order of 15%, compared to the 10^{-6} amplitude of individual eigenmodes (the ratio is about 1.5×10^5). The r.m.s. granular radial velocity can be estimated at about 1 km/s, compared to the 5 cm/s mean amplitude of individual eigenmodes (the ratio is about 2×10^4). This simple arithmetic shows that a background level significantly higher in brightness measurements is not surprising. It also indicates that it will be more difficult, in broad band photometry, to detect the lower order p-modes or the g-modes. These lie in the lower frequency part of the power spectrum, with less amplitude and more background.

4 - Atmospheric Noise.

4 - 1 - Broad Band Photometry. I have made some investigations of the ultimate sensitivity achieved in broad band photometry when facing the particular problem of searching for periodicities in the range of minutes to hours. In this investigation, I used published and unpublished observations made at La Silla, Mauna Kea, Canarie Islands, Granada and Sutherland. A somewhat surprising result is that there is very little scatter. It appears that the atmospheric limitation is roughly the same in all good mid-latitude, high altitude sites. The typical curve shown in Figure 7 gives the mean power of atmospheric noise in the case of a standard photometric telescope (aperture between 0.5 and 1 meter) and a complete night of integration with a single-channel photometer. This figure requires several important comments:

- First, how to read it: Let's take an example by saying that a good photometric night lasts about 8 hours, and consequently provides a frequency resolution of the order of 35 μHz. In the power spectrum, the noise per frequency bin at 3 mHz is in this case $35 \times 10^{-3} \times 3 \times 10^{-7} \sim 10^{-8}$. In amplitude, the r.m.s. noise is consequently of the order of 10^{-4} magnitude, or 3×10^{-4} at the $3 - \sigma$ level. This is quite consistent with the numbers reported in these proceeding by D. Kurtz. Many observers also report a noise of a few millimags per 20 , or so, seconds of integration. Is this consistent with our typical curve? 20 seconds of integration mean a 50 mHz bandwidth, in which the main noise contribution is the scintillation, at a level of about 2×10^{-7} per mHz. The noise in this band is then $50 \times 2 \times 10^{-7}$, or 10^{-5} in power. This corresponds to a r.m.s. amplitude of 3 millimags, which is indeed the typical scatter of the individual points, for example in the many good light curves of A_p stars published by D. Kurtz (1982).

- The noise level can sometimes be lower when the star travels close to the zenith; but it will not keep this low level all night long. When a long time of integration is required, a good site will typically not give better results than shown in Figure 7.

- The major contribution of the noise is the sky transparency for periods longer than 5 minutes, and the scintillation for shorter periods. This last contribution can be reduced by increasing the telescope aperture (due to the small horizontal coherence of the atmospheric inhomogeneities responsibles for the scintillation). However, this kind of observing program mostly requires a dedicated instrument, and I do not see much hope for obtaining a dedicated 4 or 5 meter telescope just to decrease the noise level by a factor of the order of $\sqrt{2}$.

- It is of major importance here to note the scale of the noise spectrum and to compare it with the scale shown in Figure 6 and corresponding to solar oscillations.

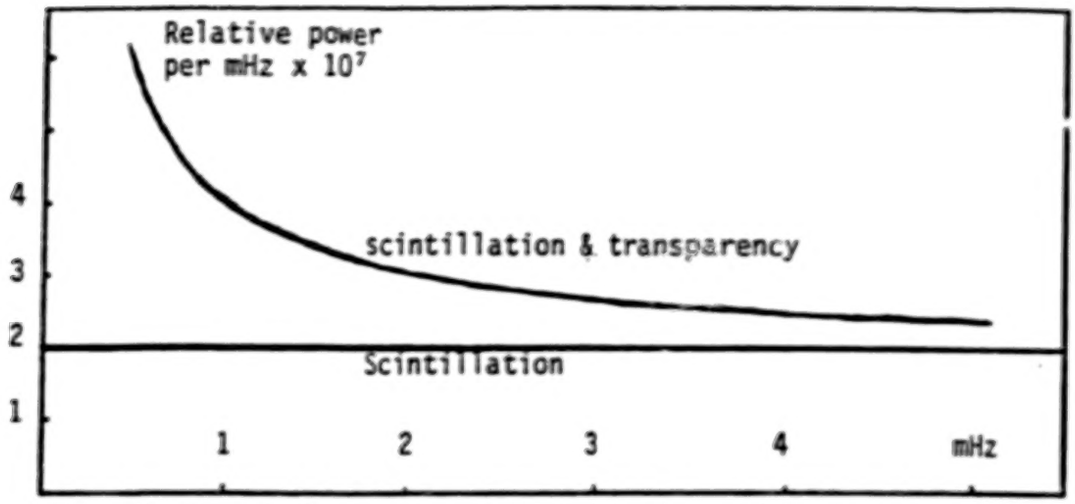


Figure 7: Typical power spectrum of atmospheric noise in all good photometric sites. Transparency fluctuations are prevailing below 3 mHz, while scintillation is the main contribution at higher frequencies.

The ratio is at least 200 in favor of the atmospheric noise with a peak width of about 1 μHz (that is in this case the real natural width defined by the lifetime of solar oscillations, and is not at all limited by the resolution of observations).

- One must then discuss how more integration can improve this situation. Two different effects are to be considered.

By increasing the duration T of the data sample, the amount of noise, in power, in the frequency interval defined by the resolution 1/T decreases as 1/T. In this case, if the signal contains a pure sine wave, the power spectrum of which is a Dirac function, the signal to noise ratio in the power spectrum is improved as 1/T. The level of amplitude detection is then improved as $\sqrt{1/T}$. However, if the power spectrum of the signal contains a broad peak, corresponding to an oscillation of limited coherence time (this is the case in the Sun, where the coherence time of the five-minute p-modes is of the order of a few days), it is useless to increase the frequency resolution beyond the threshold defined by the width of this broad peak. No more improvement of the signal to noise ratio will be obtained.

In this last case, one can limit the duration of each sample of data to the minimum which provides the adequate resolution, and then one can average all individual power spectra. (It is equivalent to smooth the unique rax spectrum by convolution of any kind of window having the here aboved called adequate resolution). The signal to noise ratio, in power, is then improved as $\sqrt{1/T}$, and the level of amplitude detection is improved only as $\sqrt[4]{1/T}$.

The solar five minute p-modes can be used to illustrate this discussion. Their mean amplitude as well as the mean width of the corresponding spectral peaks are known (Figure 6). The mean contribution of atmospheric noise at a five minute period is also known (Figure 7). Let us now assume that it is possible to obtain

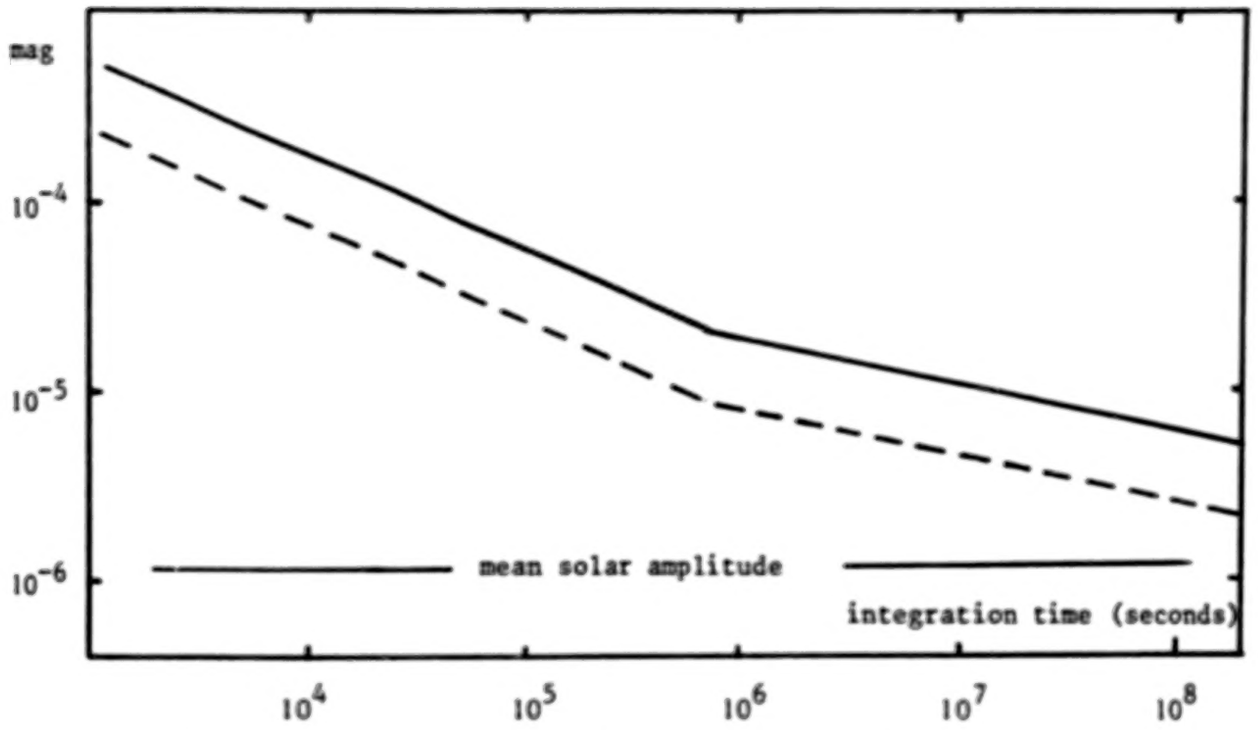


Figure 8: Detectable amplitude of a solar-like five-minute p-mode, (having a coherence time of about one week), as a function of integration time, assuming a continuous ground based observation of optimum photometric quality (full line). Using a superposed frequency analysis as in Figure 6 can improve this amplitude sensitivity by a factor 2 to 3 (dashed line).

a continuous top level photometric observation of an identical star as long as we need (for example with a conveniently located multisite network of high quality photometric instruments). The Figure 8 gives the $1 - \sigma$ level of amplitude detection as a function of the time of integration. In fact, it is possible to use the knowledge of the uniform discrete nature of the investigated power spectrum. With a superposed frequency analysis like those of the Figures 5 and 6, the $1 - \sigma$ level of amplitude detection can be improved by a factor 2 to 3, and it is given by the dashed line in Figure 8.

- In any case, even using the advantage of a superposed frequency analysis, the detection of solar-like p-modes from the ground would require at least several years of integration of top level photometric data, and that is clearly not possible.

- Now, it is quite possible that another star will display some more amplitude for similar eigenmodes. The amplitude also increases by moving the filtered band in the far blue or near U.V. If the overall gain can be as large as a factor 10 (and that seems reasonable for dwarf stars F0 to F5), the power in a peak becomes comparable to the mean noise power. This means that a week of integration time (order of magnitude of the coherence time) becomes long enough for detection.

This is compatible with the statistical weather reports at the Geographic South Pole. On the other hand, this remote site could very well be of better photometric quality than any other one. To date, nothing is known about the scintillation down there. However, my personal experience indicates that the sky transparency fluctuations are lower, perhaps by a factor 2. My conclusion on this point is that if sometime a dedicated photometric telescope is run at the South Pole, (where long continuous time series can be achieved), it should include the search for p-modes in F dwarf stars in its program.

- A more complete discussion should include the use of the two or more channel photometers to reduce the sky transparency noise. I will just say that in this particular program, it does not help. In the five-to-ten-minute range, the transparency fluctuations are not 100% correlated at the typical angular distances used in such photometers (see Grec et al, 1979), and consequently only a fraction of transparency changes are eliminated with a reference star. Moreover, the small possible decrease of sky transparency noise is balanced by an increase of scintillation noise, which is 100% uncorrelated on the two channels.

4 - 2 - Doppler Shift Measurements. Typically, such a measurement requires one (or more) monochromatic channel on the wing of a stellar line, and can use the neighbour continuum as a reference. If the same optical beam is used synchronously for the monochromatic and the reference channels, both scintillation and tranparency can be corrected and consequently, the atmospheric noise situation is much better in this case. It is even better than in the solar case because in Doppler shift solar full disk observations, a significant amount of atmospheric noise comes from the fluctuating transparency gradient in front of the rotating sun. This effect is absent in a stellar observation. It can then be photon statistics limited and this limitation depends on the magnitude of the star, the telescope aperture, and the photon efficiency of the spectrophotometer. As an illustration,

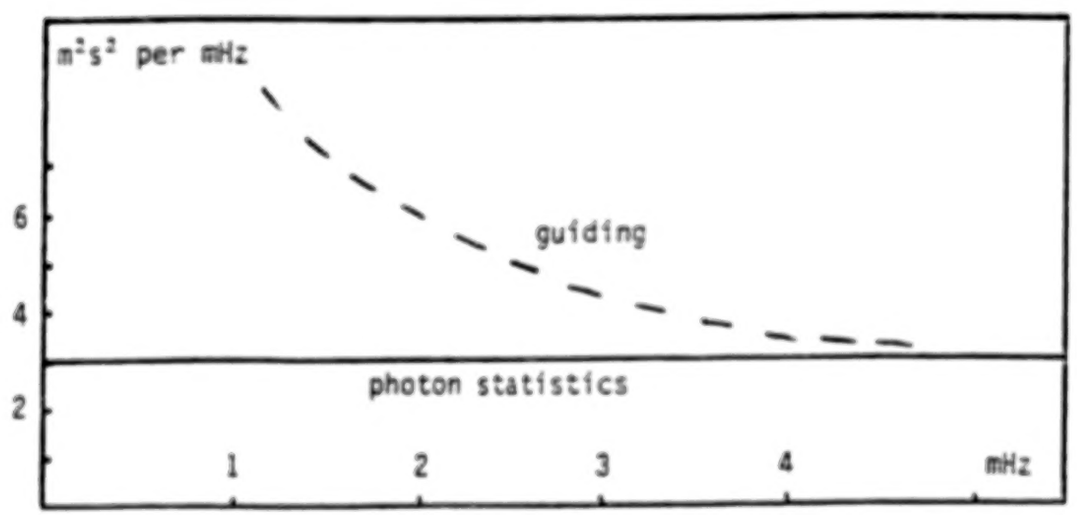


Figure 9: Example of sensitivity limitation with a spectrophotometer measuring stellar Doppler shift in a 0.1 Å bandwidth on a zero magnitude star with a 3.6 meter telescope. All atmospheric noise can be reduced at a level well below the photon statistical noise.

Figure 9 gives the result of our experience with a sodium cell monochromatic filter observing both D1 and D2 lines in 0.1 Å bandwidths, a zero magnitude star and a 3.6 meter telescope. So far the ideal level of photon noise has not completely be reached, and there is some remaining noise due to the telescope guiding fluctuations in the presence of poor seeing (we are using small aperture, due to the high price of calcite polarizers). Here again, the scale must be compared with the one in Figure 6 to note that they are very much of the same order. The success seems then possible for the detection of p-modes on very bright stars with very large telescopes, and this is confirmed in my other paper in these proceedings. The number of stars accessible to this search is very limited, but the scientific return is so evident that it is worth trying.

What can be improved in order to do more than just the detection of p-modes on a few bright stars? Our filter has a high transmission and a pretty good stability and it is very unlikely that much better will be done with any conventional instrument using one or two spectral lines. It is absolutely useless in this case to go to space because ground based observations can fairly easily be already photon noise limited. The only possible improvement that I foresee is to develop a special instrument like the one proposed by P. Connes (1984) and which makes use of the entire spectrum in an optimized manner. Following Connes, such an instrument should be capable of providing on many (magnitude ≤ 10) slow rotating stars, an accuracy comparable to the one already obtained on the Sun.

5 - Conclusions. I shall try to be very precise concerning the main points developed in this paper:

- With the existing equipment, the detection of p-modes of the brightest solar-like stars is possible from the ground in Doppler shift measurements. It is not possible in broad band photometry. However, moving to dwarf F0 to F5 stars, this detection is presumably possible in broad band blue or UV photometry, with a dedicated telescope operated at the South Pole.

- The new accelerometer proposed by P. Connes would, in principle, make possible to obtain from the ground the precision which has already been obtained on th Sun, and this for many slow rotating stars of the magnitude $m_v < 10$. The instrument does not exist and it may prove to be difficult to realize but meanwhile, it seems that a network of two coordinated such accelerometers could provide for many stars the power spectrum obtained for the Sun in Figure 4, with even better quality in the low frequencies.

- To avoid being limited to slow rotating dwarf stars, broad band photometry must be made in space. There, a more extensive investigation of the dynamical interaction between convection, rotation and oscillations will be possible. Possibly also, the accuracy of the measurement will be limited by the star itself through its convective background noise in the low frequency range. But this depends certainly very much on the rotation speed, and it will be an information by itself.

References

Solar Physics 1983, 82 (Proceedings of the Crimean Conference on Problems of Solar and Stellar Oscillations, September, 1981); all papers

Proceedings of the EPS Study Conference held in Catania, May, 1983; L. Parteno and G. Belvedere, eds

Proceedings of the Snowmass Conference on Solar Seismology from Space; Snowmass, August, 1983; R.K. Ulrich, ed.

CONNES, J.: 1984, Proceedings of the Workshop on Space Prospects in Stellar activity and variability; Meudon, March, 1984; F. Praderie, ed.

CONNES, P.: 1984, Same Proceedings

FOSSAT, E.; DECANINI, Y. and GREC, G.: 1982, in Instrumentation for Astronomy with Larger Telescopes, September, 1981; C. Humphries, ed.

GREC, G.; FOSSAT, E.; BRANDT, P.; and DEUBNER, F.L.: 1979, Astron. Astrophys. 77, 347

KURTZ, D.: 1982, Mont. Not. R. Astr. Soc. 200, 807

WOODARD, M.; and HUDSON, H.: 1983, Nature 305, 589

N85-17900

A STATISTICAL EVALUATION OF THE LIMITATIONS OF SINGLE-CHANNEL INTERMEDIATE-BAND PHOTOELECTRIC STELLAR PHOTOMETRY

G. W. Lockwood
Lowell Observatory

Introduction. We have had a decade of experience in the differential photometry of stars, the outer planets Uranus and Neptune, and the brighter satellites of Jupiter and Saturn, pursuant to the question of long-term solar variability. (A cynical view is that we have one year's experience ten times!) Our goal is to maintain long-term accuracy to better than 0.3% for a moving solar system object by comparing it each year to a pair of nearby comparison stars roughly of solar color. A parallel program measures the relative magnitudes of the annual pairs of comparison stars in a quasi-differential mode.

Various aspects of the technique and results are given in Lockwood (1977, 1983) and in Lockwood, Thompson, and Lumme (1980). Our claim that systematic and random errors are understood and controlled at the level of ~0.3% is supported by a light curve of Europa and Callisto combined, in which the rms scatter of the annual mean magnitude is 0.14%; hence the 95% confidence interval for the seven-year series is ±0.3%. The slope of the relation is -0.03% yr^{-1}, at the statistically insignificant confidence level of about 80%. Perhaps fortuitously, this apparent slope is in agreement with the output of radiometers on the Solar Maximum Mission and Nimbus satellites.

In this paper, we review the results of differential photometry of several dozen pairs of planetary comparison stars observed since 1972 by myself and D. T. Thompson, in a program supported by the Division of Atmospheric Sciences of the National Science Foundation. The statistical tests described herein were carried out using the Minitab™ statistical computing system.

Each pair of stars was observed along with the corresponding planet or satellite, typically about ten times during the course of a single apparition lasting about four months. The comparison stars are chosen to bracket the opposition position of the planet, the expected opposition magnitude, and solar color. In practice, this means that the range of spectral type was mid-F to early K, the range of b-y color was about 0.25 to 0.95 mag, the range of differential magnitude was less than 2 mag but most often less than 1.0 mag, and the difference in declination was typically less than 3°. The difference in air mass at meridian transit was usually less than 0.03, and rarely exceeded 0.1. Hence, differential extinction effects are negligible, except under extraordinary conditions as described below.

Characteristics of the Instrumentation. The data obtained in this experiment are perhaps unique in their homogeneity owing to the long-term nature of the observing program. The same telescope, photometer, photomultiplier, interference filters, and pulse-counting electronics have been used throughout, except for rare changes as noted: The 0.5-m telescope was used from 1974 onward; prior to that (1972-1973), larger telescopes were used. In 1982, the NIM-type pulse-amplifier/discriminator modules were replaced by a Pacific Photometric model AD-6 amplifier/discriminator. In 1983, the DEC PDP-11/05 data acquisition mini-computer and its associated interface electronics, including the pulse counter, were replaced by a DEC LSI-11/03

microcomputer with associated ADAC interface modules, including a high-speed pulse counter, and flexible disk drive. From the beginning, the EMI type 6256S photomultiplier has been housed in an early-1970s vintage Products for Research thermoelectric cooler, and operated at -15°C and 1.3kV. The photometer contains, in addition to the b and y filters, an Sr^{90} Cerenkov light source, measured unfiltered, whose output is recorded as a routine part of all observations. However, these data are unused in the course of routine stellar data reduction; they are recorded only as a long-term record of system sensitivity and stability.

Observational Procedure. Precisely the same observational procedure, originally implemented by M. Jerzykiewicz in 1972, has been used for all of the observations, with only occasional deviations. A "cycle" of differential photometry through a given filter (b or y) proceeds as objects 2-1-1-3-3-2, where 1, 2, and 3 refer to the planet and its two comparison stars, respectively. All integrations are 10 seconds long and are taken in the order

2sk, 2sk, 2st, 2st, 2st
1st, 1st, 1st, 1sk, 1sk, 1st, 1st, 1st
3st, 3st, 3st, 3sk, 3sk, 3st, 3st, 3st
2st, 2st, 2st, 2sk, 2sk,

where "st" indicates an observation of the star, and "sk" indicates the sky. The various differential magnitudes taken pairwise (e.g., 2 minus 1) have the same weight with respect to scintillation noise, etc. In practice, the planetary differential magnitudes are the average of 1-2 and 1-3, while in this discussion we consider exclusively the differential magnitudes of the comparison stars 2 and 3, i.e., 2-3. Poisson statistics are negligible at levels ranging from 0.0008 to 0.0014 mag over the range of brightness considered here (V = 5 to 8.5 mag); hence in the discussion which follows we need concern ourselves only with instrumental, atmospheric, and (perhaps) stellar effects.

The nightly values have equal statistical weight, since with rare exception each comprises four cycles in y and four in b, taken in the order y-y-b-b-b-b-y-y, within an hour of meridian transit. A season's data thus consist of a sequence of nights, beginning with morning observations and ending four months later with evening observations.

Possible Instrumental Effects. Observations are often begun immediately after opening the roll-off roof; during the first hour, the ambient temperature drops about 5°C and continues to drop more slowly throughout the night. The total diurnal range is frequently 15 or 20°C, as is typical of dry sites at this elevation (2.2 km).

The overall photometer sensitivity, as measured by the standard source, increases by about 1% during the first hour of operation as the temperature falls and exhibits, in addition, a seasonal range (winter to summer) of about 10%. We don't know the cause, but we can imagine several possibilities: Although the amplifier-discriminator and high-voltage power supply are always left on day and night, they operate at ambient temperature, ameliorated only by their internal heat dissipation. A temperature effect in overall system gain is therefore plausible. Less plausible is a change in photomultiplier cathode temperature, since its temperature is regulated on the basis of a probe inside the chamber.

Experiments involving sequential measurements of the standard source at constant outside ambient temperature reveal absolutely no deviations from Poisson statistics at the expected level of 0.001 mag over several hours. Power spectra of

these data show no harmonic content, such as might be imagined from possible thermal cycling of the coldbox. Hence, for the purpose of a differential measurement lasting for a hour or so, instrumental sensitivity changes by negligible amounts. The observed temperature coefficient, determined from changing the internal temperature from 0°C to -15°C is about +0.001 mag/°C at a constant outside ambient temperature of about 10°C. The observed seasonal (winter to summer) range is about 0.1 mag, which implies a temperature coefficient on the order of +0.005 mag/°C. This discrepancy needs to be resolved but, in any case, does not seem to be relevant to the present discussion.

The temperature of the interference filters is not regulated, and thus tracks ambient temperature. However, we have made tests of the night-to-night transformation coefficients for standard star observations and of raw instrumental differential magnitudes of comparison stars and found no evidence of temperature effects. Further investigation amd experiments related to this particular question are intended.

A Fundamental Problem. We take a statistical approach in describing the propagation of observational error in a series of differential observations, beginning with the repeatability during intervals shorter than one minute and ending with the repeatability from night to night. A fundamental problem whose complete solution still eludes us is the following: we find consistently that the cycle-to-cycle rms scatter of differential magnitudes on a given night is typically 0.004 mag, which suggests that a nightly series of four cyles should have an internal precision of 0.002 mag. With rare exception, we find that the night-to-night repetition over a series of nights is on the order of 0.003 to 0.004 mag, which suggests that three-quarters of the night-to-night variance is of unknown origin, although we argue below that extraordinary atmospheric conditions may be the cause. A series of statistical experiments to be described addresses this question, which is fundamental with regard to any photometric program whose goal is a long-term relative accuracy on the order of 10^{-3} (0.001 mag).

The Precision of a Single Cycle. A standard cycle of observation 2-1-1-3-3-2 takes less than 8 minutes, of which about half is consumed by measurement and half in setting the telescope, checking centering, etc. Are changes of sky transparency important on this time scale? We examine this question by considering the dispersion of the differential magnitudes of star 2 compared with itself (separated by about 7 minutes of time), star 3 compared with itself (separated by less than 1 minute), star 2 compared with star 3 (separated by about 1 minute, i.e., the adjacent pair listed in the standard sequence), and star 2 compared with star 3 (separated by about 5 minutes, i.e., the observations bracketing the planetary object 1). Table I describes the observed distribution of differential magnitudes of Uranus comparison stars obtained in 116 cycles on 29 nights in 1978, and 52 cycles on 13 nights in 1977. The declination was about -15°.

We chose to use the median and quartiles to describe the distribution of differential magnitudes in these two examples because the distributions are typically skewed. Note that the ranges of quartile and median values are essentially independent of the time interval, which ranges from less than a minute to more than 7 minutes. We conclude that temporal changes of extinction are unimportant on this time scale. Hence, rapid chopping between stars is not expected to yield significant improvements in differential photometry. The distributions of differential magnitudes are not normal but are skewed toward higher values with means about 25% larger than the medians. Representative histograms of differential magnitudes from the data in Table I are shown as Figure 1.

TABLE I. An Example of Repetition on Short Time Scales.

Year	1977	1978
Number of cycles/nights	52/13	116/29
Star 2-star 2(Δt ~ 7 min)	0^m.0021 .0036 .0065*	0^m.0014 .0037 .0056
Star 2-star 3(Δt ~ $4\frac{1}{2}$ min)	.0016 .0033 .0069	.0024 .0044 .0075
Star 2-star 3(Δt ~ 1 min)	.0023 .0043 .0066	.0017 .0039 .0066
Star 3-star 3(Δt < 1 min)	.0021 .0035 .0052	.0014 .0029 .0049
Avg σ (cycle-to-cycle)	.0034	.0034
Expected night-to-night repetition (rms)	.0017	.0017
Observed night-to-night repetition (rms)	.0032	.0044

*(First quartile, median, third quartile) for each entry.

STAR 2 - STAR 2 1977

52 cycles in 13 nights

Middle of Interval		Number of Observations	
0.0000		3	***
0.0005		2	**
0.0010		4	****
0.0015		2	**
0.0020	Q1	4	****
0.0025		3	***
0.0030		6	******
0.0035	M	3	***
0.0040		2	**
0.0045	Mean	1	*
0.0050		6	******
0.0055		2	**
0.0060		0	
0.0065	Q3	2	**
0.0070		1	*
0.0075		3	***
0.0080		0	
0.0085		1	*
0.0090		0	
0.0095		0	
0.0100		3	***

(plus four more values to 0.0195)

STAR 2 - STAR 2 1978

116 cycles in 29 nights

Middle of Interval		Number of Observations	
0.000		5	*****
0.0005		12	************
0.0010		8	********
0.0015	Q1	6	******
0.0020		7	*******
0.0025		8	********
0.0030		4	****
0.0035	M	9	*********
0.0040		8	********
0.0045	Mean	7	*******
0.0050		11	***********
0.0055	Q3	3	***
0.0060		2	**
0.0065		6	******
0.0070		4	****
0.0075		2	**
0.0080		1	*
0.0085		2	**
0.0090		2	**
0.0095		1	*
0.0100		0	

(plus eight more values to 0.0185)

Figure 1. Histograms of the differential magnitudes of star 2 compared with itself in two representative data sets described in Table I. Q1, M, and Q3 indicate the approximate locations of the first quartile, median, and third quartile.

We see in Table I that the star-to-star repetition is quite consistent with the cycle-to-cycle repetition of 0.0034 mag for two different pairs of stars in two seasons. The expected night-to-night repeatability is better than 0.002 on the basis of the precision of the nightly observation; yet we observe substantially larger values (bottom line of the table). This phenomenon is quite typical, in fact universal, which implies that (1) all stars are variable or (2) we have an unknown source of variance amounting to approximately $(0.004^2 - 0.002^2)^{1/2} = 0.003$.

The Distribution of Cycle-to-Cycle Variations. The average (over a season) cycle-to-cycle variations of various pairs of stars are quite significantly different from one another for no obvious reason. Figure 2 illustrates the distribution of observed cycle-to-cycle variations for 34 pairs of stars observed on at least four nights each. The mean dispersion in y is significantly larger than in b, by about 10%, and the dispersions are highly correlated. An analysis of variance of groups of stars associated with three different planetary objects having the characteristics shown in Table II reveals no significant differences despite a wide range of average brightness, photometer diaphragm size, etc. Further investigation is needed to understand why we get better results for some pairs of stars than for others. A possibility to be considered is the contamination of sky background measurements by faint stars, whose average color would lead to larger effects in y than in b. Our archived raw data files will be used to investigate the question.

y Filter			b Filter		
Middle of Interval	Number of Observations		Middle of Interval	Number of Observations	
0.0000	0		0.0000	0	
0.0005	0		0.0005	0	
0.0010	0		0.0010	0	
0.0015	0		0.0015	0	
0.0020	0		0.0020	0	
0.0025	1	*	0.0025	3	***
0.0030	2	**	0.0030	3	***
0.0035	3	***	0.0035	5	*****
0.0040	5	*****	0.0040	7	*******
0.0045	6	******	0.0045	4	****
0.0050	7	*******	0.0050	6	******
0.0055	4	****	0.0055	5	*****
0.0060	1	*	0.0060	0	
0.0065	0		0.0065	0	
0.0070	2	**	0.0070	1	*
0.0075	2	**			
0.0080	1	*			

Figure 2. Histograms of mean rms cycle-to-cycle dispersions of differential magnitudes of 34 pairs of stars, averaged over all nights.

TABLE II. Average properties of three subgroups of comparison stars whose cycle-to-cycle variations are illustrated in Figure 2.

	Uranus Comparison Stars	Neptune Comparison Stars	Titan Comparison Stars
Number of pairs	10	12	12
Mean of y mag	6.5	8.0	8.2
Star diaphragm size (arcsec)	29-49	49	19
Mean dispersion in y	0.0046	0.0047	0.0053
Mean dispersion in b	0.0038	0.0046	0.0044

The Distribution of Night-to-Night Variability. We have examined the characteristics of the night-to-night variability of 44 pairs of comparison stars to see if there are discernible differences due to magnitude, color (spectral type), magnitude difference, color difference, mean air mass, air mass difference, and various circumstances of the observations. Typically, the data were divided into subgroups according to the range of some parameter (e.g., y magnitude), and differences between subgroups were investigated using a one-way analysis of variance. These tests are summarized below.

1. **Correlations of Δb and Δy.** The rms dispersions in b and y are highly correlated, as one expects. The mean dispersion in y is slightly smaller than in b, but the medians are identical, as Table III shows.

TABLE III. Statistics of Night-to-Night Variability.

	y	b
Number of pairs	44	44
Mean absolute deviation	0.0043	0.0045
Median	0.0036	0.0036
First quartile	0.0027	0.0026
Third quartile	0.0055	0.0059

Histograms of the seasonal dispersions of nightly differential magnitudes are shown in Figure 3.

rms Dispersion in *y*			rms Dispersion in *b*		
Middle of Interval	Number of Observations		Middle of Interval	Number of Observations	
0.0000	0		0.0000	0	
0.0005	0		0.0005	0	
0.0010	0		0.0010	0	
0.0015	2	**	0.0015	3	***
0.0020	4	****	0.0020	3	***
0.0025	6	******	0.0025	6	******
0.0030	6	******	0.0030	5	*****
0.0035	6	******	0.0035	6	******
0.0040	1	*	0.0040	2	**
0.0045	2	**	0.0045	2	**
0.0050	4	****	0.0050	2	**
0.0055	4	****	0.0055	2	**
0.0060	1	*	0.0060	4	****
0.0065	1	*	0.0065	1	*
0.0070	3	***	0.0070	1	*
0.0075	0		0.0075	2	**
(plus three values up to 0.11)			(plus five more values up to 0.013)		

Figure 3. Distribution of the night-to-night dispersions of 44 pairs of comparison stars, spectral types F, G, and early K.

2. **Spectral type.** The stars were grouped according to the color of the "blue" star, the color of the "red" star, and the color difference within the pair. No significant effects were found, which indicates no tendency for variability to be associated with spectral type or spectral type difference. This test specifically addresses the question of possible instrumental effects associated with temperature-dependent shifts in the *b* and *y* bandpasses.

3. **Magnitude and magnitude difference.** The stars were grouped according to the brightness of the fainter member of the pair and the magnitude difference between the pair. No significant effects were found. This test confirms the linearity of the photometer, the absence of dead-time errors at the bright end, and the absence of perceptible photon noise at the faint end.

4. **Effects of the declination zone, and the difference in air mass.** Stars were grouped according to declination (in the range +20° to -23°) and according to the difference in air mass (in the range from 0.0 to about 0.1). No significant effects were found, which, surprisingly, indicates that our photometry at air mass 2.0 is no worse than at air mass 1.0.

5. **Effects due to moonlight, seeing, diaphragm size, etc.** We tested for effects associated with the presence or absence of moonlight, the quality of the seeing, the diaphragm size used, and the season of the year. No significant effects were found. A specific conclusion is therefore that "winter" objects, observed in the ecliptic at +20° declination under conditions of relatively low air mass and seasonally low extinction do not yield results significantly different from those obtained on "summer" nights at -23° declination, high air mass, and high seasonal extinction.

Effects of Abnormal Extinction and Light Cirrus. We have previously adopted the conventional wisdom that differential extinction effects are negligible for small (<0.1) air mass differences, but we show evidence here that this assumption is sometimes invalid. All observers note that some nights are better than others even in the apparent absence of daytime or nighttime clouds. Occasionally thin cirrus is seen on the Flagstaff horizon, with no immediately evident ill effects upon the night's work.

However, we have long noted, without much concern, that the count rate varies perceptibly from night to night, even though the repetition of measurements over an hour or two seems quite good. The stability of our photometer over long time periods (weeks) allows a retrospective analysis of night-to-night variations, which from winter to summer typically have a range of about 0.2 mag or more. The instrumental sensitivity (raw count rate) is known to be sensitive to temperature, so this was no surprise. Further, the extinction increases from winter to summer, leading to additional changes in the same direction. To search for possible extinction or instrumental effects leading to systematic errors in differential magnitudes, we have carried out an analysis of the differential photometry of seven pairs of Uranus comparison stars observed in seven different years.

The results of the preliminary analysis are summarized in Table IV, where the top section gives the relevant properties of the particular pairs of stars considered. Here, σ is the rms scatter of the n nights of observation, Δz is the average air mass difference. In each section below, r gives the significance of correlation between the differential magnitude Δb and b, the raw blue magnitude of one of the stars corrected only for mean extinction, y^*, the raw yellow magnitude normalized by the raw count rate of the standard source, the raw b-y color, and finally by the raw count rate of the standard source itself. The subsequent entries give "σ fit," the estimated standard deviation about the regression line, the slope if significant at the 95% level, and the percentage of the total variation which is explained by the regression.

In about half of the cases, the scatter (σ fit) is significantly improved by regression relative to the value in line 4 of the table. In the three years with $\Delta z > 0.05$, regression against b, y^*, or std leads to improvement, which suggests that there may be systematic differential extinction effects, winter to summer, which are also, of course, correlated with the known temperature effects. In two seasons with $\Delta z \sim 0$, no significant correlations exist. Three seasons show significant correlations with raw b-y, which suggests the presence of a small color error either in the extinction or the instrument. It may be appropriate to attribute some of these effects to light cirrus, which increases the extinction by a substantial amount.

Conclusions. We have described the observed distribution of errors in differential photometry on time scales from minutes to months, based on a large and unusually homogeneous data set. Four cycles each of b and y differential observations taken in the course of an hour at meridian transits yield a fairly predictable internal precision of 0.2% (0.002 mag). The night-to-night repetition is significantly worse, by at least a factor of 1.5, leading to the suspicion of very low-level systematic instrumental or atmospheric effects. The possibility of intrinsic stellar variability cannot be ruled out, although a preliminary statistical analysis of a few data sets reveals evidence for effects amenable to further analysis. These will be addressed in a forthcoming publication.

TABLE IV. Some experiments with comparison star data.

Year	1975	1976	1977	1978	1979	1980	1981
n	11	9	13	28	22	11	11
$\Delta\bar{b}$	+0.46	+0.43	+0.44	-1.53	-1.98	+0.91	+1.06
$\Delta(\bar{b}-\bar{y})$	+0.27	-0.23	0.00	-0.36	+0.67	-0.46	-0.23
$\Delta\bar{z}$	0.12	0.01	0.06	0.10	0.02	0.13	0.02
σ	0.0039	0.0048	0.0032	0.0036	0.0042	0.0055	0.0014
$r(\Delta\bar{b},\bar{b})$	99.5%	-	99.9%	99.9%	-	-	-
σ fit	0.0025	-	0.031	0.057	-	-	-
slope	0.065	-	0.031	0.057	-	-	-
r^2	61%	4%	62%	55%	0%	25%	0%
$r(\Delta\bar{b},\bar{y}^*)$	99.5%	-	-	-	-	-	-
σ fit	0.0025	-	-	-	-	-	-
slope	0.078	-	-	-	-	-	-
r^2	59%	0%	2%	0%	0%	9%	10%
$r(\Delta\bar{b},\bar{b}-\bar{y})$	-	95%	99.5%	-	-	99%	-
σ fit	-	0.0036	0.0022	-	-	0.0039	-
slope	-	0.190	0.112	-	-	0.262	-
r^2	0%	44%	54%	9%	8%	50%	0%
r(Δb̄,std)	-	-95%	95%	95%	-	-	-
σ fit	-	0.0035	0.0026	0.0034	-	-	-
slope	-	-1.013	0.031	0.022	-	-	-
r^2	13%	47%	34%	15%	0%	0%	17%

References

1. Lockwood, G. W.: Secular Brightness Increases of Titan, Uranus, and Neptune, 1972-76. Icarus, vol. 32, 1977, p. 413.

2. Lockwood, G. W.: Photometry of Planets and Satellites, in Solar System Photometry Handbook, R. M. Genet, Ed., Willmann-Bell, Inc., 1983, pp. 2-1 through 2-19.

3. Lockwood, G. W.; Thompson, D. T.; and Lumme, K.: A Possible Detection of Solar Variability: Photometry of Io, Europa, Callisto, and Rhea, 1976-1979. Astron. J., vol. 85, 1980, p. 961.

THE ATMOSPHERIC EXTINCTION PROBLEM

Ronald J. Angione
Astronomy Department
San Diego State University

Atmospheric extinction is one of the main causes of errors in photometry. The incorrect determination of the extinction coefficient, and its variability, leads to an erroneous measurement.

An erroneous extinction coefficient can arise from a number of causes including (1) instrumental instabilities, (2) too few data points, (3) temporal changes in the atmosphere, (4) differing airmasses due to components with different scale heights. While it is true that differential measuring techniques can achieve a precision approaching 0.1 percent, at higher levels of precision all of the above causes will be significant sources of error. The whole area of milli and micro-extinction, and rapid, small amplitude changes (time-scale less than one second), is still unexplored. For example, consider two close stars differing in airmass by 0.01, an error of 0.01 in the extinction coefficient causes an error of 0.0001. Extinction coefficients are rarely determined to ±0.01 magnitudes/airmass. Nor with an assemblage of stars can you keep the differential airmass as small as 0.01, thus the demands on the accuracy of the extinction coefficient will be at least ± 0.001 mag./airmass.

One must begin with a highly stable instrument, otherwise high precision is hopeless. Our engineer, Frank Beale, has designed and built a very stable photometer for solar measurements. The optical design is similar to a stellar photometer. Sunlight enters through a 5 mm aperture behind a quartz window. A quartz Fabry lens images this aperture on a EG&G UV 444B silicon photodiode detector. A twelve position filter wheel lies between the Fabry lens and the detector. The diode is temperature controlled to + 0.1 ^{o}C, and the entire photometer is controlled to ± 0.5 ^{o}C year round. Much of the electronics are inside the photometer box and are thus kept at a constant temperature also. The photometer will be shortly kept under a positive pressure of dry nitrogen to protect the filters. Because the photometer is sealed, one must worry about outgassing residues coating the optical surfaces. The radiometer is carried on an active solar tracking mount (± one arcminute, FOV = one degree). Although this is a solar radiometer, many of the design techniques are applicable to stellar photometers and are being incorporated into our new photometer the 1-meter telescope at Mt. Laguna Observatory.

Figure 1 shows a flow diagram of our measuring procedure. The radiometer has eleven interference filters (FWHM approximately 75 Å) strategically placed from 0.38 to 1.0 microns to give us ozone from the Chappuis band, water vapor from the 0.94 micron region, and the dust component from the remaining filters.

The extinction, or here really optical depth ($k = 1.086\tau$) is determined from the slope of Bouguer curves (ln(I) versus airmass). Figure 2 shows such a curve for our shortest wavelength, and Fig. 3 shows the Bouguer curve for our longest wavelength. Although the scatter may appear large in Fig. 3, note that the scale division is 0.001 mag., or 0.1%. The individual fluctuations are interpreted as

real changes in the optical depth. Although there are weak water vapor bands present, the small amplitude, short-time scale changes seen in this figure are also seen at the other water-free wavelengths, and thus these variations are not due to water vapor changes. Nor are these variations instrumental, for in that case the variations would have an amplitude independent of wavelength. In fact, the amplitude of the variations is nearly proportional to the reciprocal of the wavelength, which is also directly proportional to the aerosol component.

Figure 4 shows Bouguer curve results for night observations with our 16-inch telescope and a typical astronomical photometer (1P21 and dry ice). Note that the fluctuations are about ten times larger, and are here interpreted as due to instrumental errors. The conclusion from this is that a properly designed and built photometer can easily improve the measurement precision by an order of magnitude.

Extinction obviously changes with time, and to illustrate that change Fig. 5 (Roosen and Angione, 1984) shows the time series of optical depth obtained from Smithsonian data for Mt. Montezuma, Chile (1923 to 1955). You can readily pick out a seasonal component and the volcanic signature of the Quizapu eruption in 1932. There are, of course, many other changes also evident in this data set. The main point is that extinction is highly variable and must be accurately determined for precision photometry.

The seasonal variation is illustrated in Fig. 6 (Roosen, Angione, Klemcke, 1973). Note in particular that the seasonal variation is not well behaved sinusoid, but it is the case that in summer the amplitude of the extinction variation becomes both large and variable. There can also be large changes from one night to the next, which means that the extinction is changing during the night and day.

Time dependent extinction changes have been known for some time, and have been discussed by Young (1974). Fig. 7 shows a classic example of such a change. As the morning progresses, solar heating causes convection and evaporation, which puts more aerosols into the atmosphere; it may also be the case that the moisture causes existing aerosols to swell in size to the range where they will be optically active in the visible region. In any case, the extinction increases with time and the afternoon Bouguer curve shows a much steeper slope.

This increased afternoon extinction must then decrease during the night, suggesting that we should see a reverse effect on the night-time Bouguer curves. This is the interpretation of Fig. 8. The extinction (from the Mt. Laguna 16-inch telescope) is large at the start of the night (large airmass). As the night progresses the star anomalously brightens up as the "dust settles", or aerosols shrink. These last two figures also point out an inherent weakness of the Bouguer curve method of determining extinction - it is difficult to separate out time-dependent extinction changes except in the more obvious cases.

Volcanic eruptions also play havoc with extinction (see for example de Vaucouleurs and Angione, 1974). Because we are still in the process of accumulating data on the eruption of El Chichon, the eruption of Quizapu in 1932 will be used to characterize the effects of volcanoes on atmospheric extinction. The four curves in Fig. 9 represent the residual optical depth after the average quie-

scent monthly mean value was subtracted. Note (1)that the aerosol optical depth nearly doubles, (2) that the optical depth is quite variable, and (3) that it takes about 2 years for the stratospheric particles to settle out. At long wavelengths, top curve, the particles are larger, and hence they settle out faster.

Modelling the extinction is quite difficult. You normally don't have enough information to even calculate the Rayleigh component better than one percent. There is a lot of misunderstanding about Rayleigh scattering (see Young, 1982). It is common to assume that the aerosol component consists of particles whose sizes are distributed in a log-normal fashion ($dN(r)/d\log(r) = Cr^{-m}$). This is based in part on measurements of tropospheric particle sizes. In this case, the extinction coefficient, $k = B\,\lambda^{-n}$, with "n" about 1. Figure 10 shows such a fit.

After the eruption of El Chichon I found I could no longer represent the aerosol contribution by $k = B\,\lambda^{-n}$. There were higher order terms needed. It is also clear that "n" must be a function of wavelength because for large particles n = 0 and for small particles n goes to 4. This is also discussed by Nicholis (1984). Toon and Pollack (1976) used a Zold distribution function (based on particle sizes in colloidal suspension) to model the stratospheric dust, and they obtained a similar wavelength dependance for what we find after the El Chichon eruption; remember the volcanic dust we see is stratospheric rather than tropospheric. The real situation is more complicated because the tropospheric dust is still likely a log-normal distribution and the stratospheric dust a Zold distribution. The result may be that you can't do accurate photometry during times of large amounts of volcanic dust.

Water vapor variations are important for accurate photometry. Not only are there water vapor bands all over the visible and IR spectrum, but also increased water vapor may be absorbed onto aerosols causing them to swell and change size thus affecting the aerosol component of the extinction in addition to the absorption part. Variations in water vapor can also take place on short time scales, such as during the night. Figure 11 shows that there is a strong correlation between water vapor and extinction. Thus water vapor must be monitored. This can be done as shown in Fig. 12 by using the calibrated ratio between the intensity through a filter on the continuum and that for one in a water vapor region. Figure 13 (from Roosen and Angione, 1977) shows the time series variability of water vapor (seasonal variation has been removed) to illustrate the kinds of variations to expect from water vapor at typical mountain sites.

Figure 14 shows the extent of the Chappuis band ozone region; it is even worse in the ultraviolet. At about 6000 Anstroms the ozone accounts for about 0.03 magnitudes of extinction. Figure 15 (Angione and Roosen, 1983) shows the time series variability of ozone. It can change by up to 25% during a day or night. This means a change of 0.0075 magnitudes in the extinction. Ozone can an should be monitored during the night just as we do in the daytime in order to achieve accurate photometric measurements.

The problem of different airmasses for different atmospheric constituents has been discussed by Young (1974). Following Young (1969) the airmass M(z), where z is the zenith angle, for an exponential atmosphere is:

$$M(z) = \left(\frac{R}{2h_o}\right) \cdot \exp\left(\frac{R\cos^2 z}{2h_o}\right) \cdot \operatorname{erfc}\left(\frac{R\cos^2 z}{2h_o}\right)^{1/2}$$

where R is the Earth radius and h_o is the scale height. Even for a constant z, the airmass will clearly be different for different scale heights. Rayleigh scattering has a scale height around 8 km, aerosols about 1 km, water vapor 1 - 2 km, and ozone is concentrated in the stratosphere. Using these numbers and R = 8599 km, at a Rayleigh airmass of 1.5, we underestimate the aerosol airmass by 0.0028 airmasses; at airmass 3 the difference is ten times larger.

During times of perturbed atmospheric conditions (e.g. volcanic eruptions) the scale height problem will be worse - half the aerosol may be in the stratosphere.

The main conclusion of this paper is that the spatial and temporal variations in atmospheric extinction set a fundamental limit to the precision of ground-based photometry. In addition, the following conditions must be met in order to properly determine and correct for atmospheric extinction.

1. Use a very stable photometer.

2. Use only the best of nights, rejecting those with high extinction, high water vapor, non-repeatable measurements, ect.

3. Make frequent observations of standard stars in order to be able to detect time-variable extinction.

4. Restrict program observations to near the zenith in order to keep the extinction correction small.

5. Reduce data together in sets of 5 - 7 nights; keep the time span covered by the data set less than the time scale of the Rossby waves.

REFERENCES

Angione, R.J., and Roosen, R. G. (1983). J. Climate and Appl. Met. 22, 1377

de Vaucouleurs, G., and Angione, R. J. (1974). P.A.S.P. 86, 104.

Nicholis, R. W. (1984). Appl. Opt. 23, 1142.

Roosen, R. G., and Angione, R. J. (1977). P.A.S.P. 89, 814.

Roosen, R. G., and Angione, R. J. (1984). Bull. Amer. Met. Soc. (in press, Sept.)

Roosen, R. G., Angione, R. J., and Klemcke, C.H. (1973). Bull. Amer. Met. Soc. 54, 307.

Toon, O.B., and Pollac,, J.B. (1976). J. Appl. Met. 15, 225.

Young, A.T. (1969). Icarus 11, 1.

Young, A. T. (1974). Chapter 3, in Methods of Experimental Physics, ed. N. Carleton (Academic Press, New York).

Young, A.T. (1982). Physics Today, January, 2.

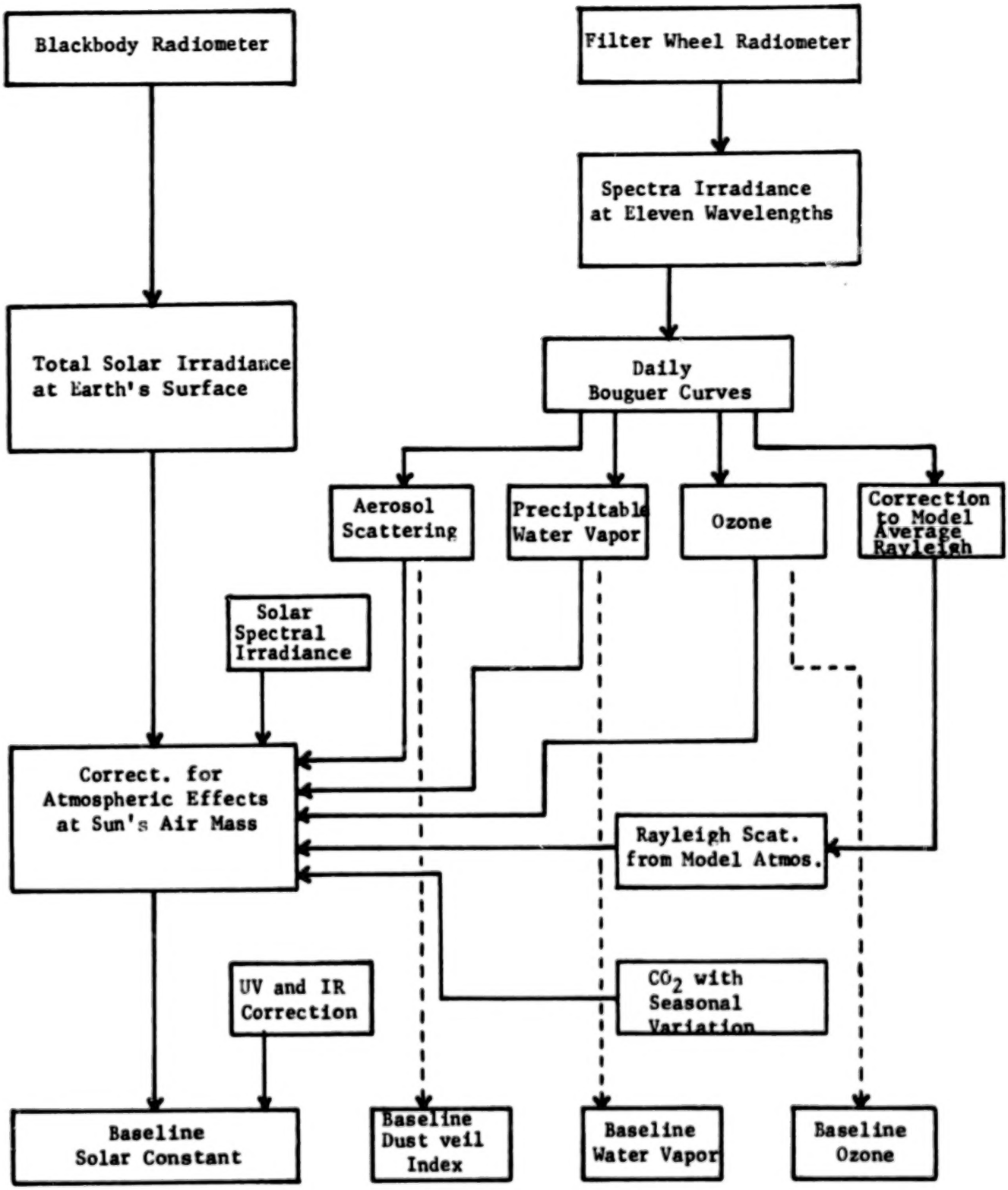


Figure 1. Flow diagram of the solar constant measurement program at Mount Laguna Observatory.

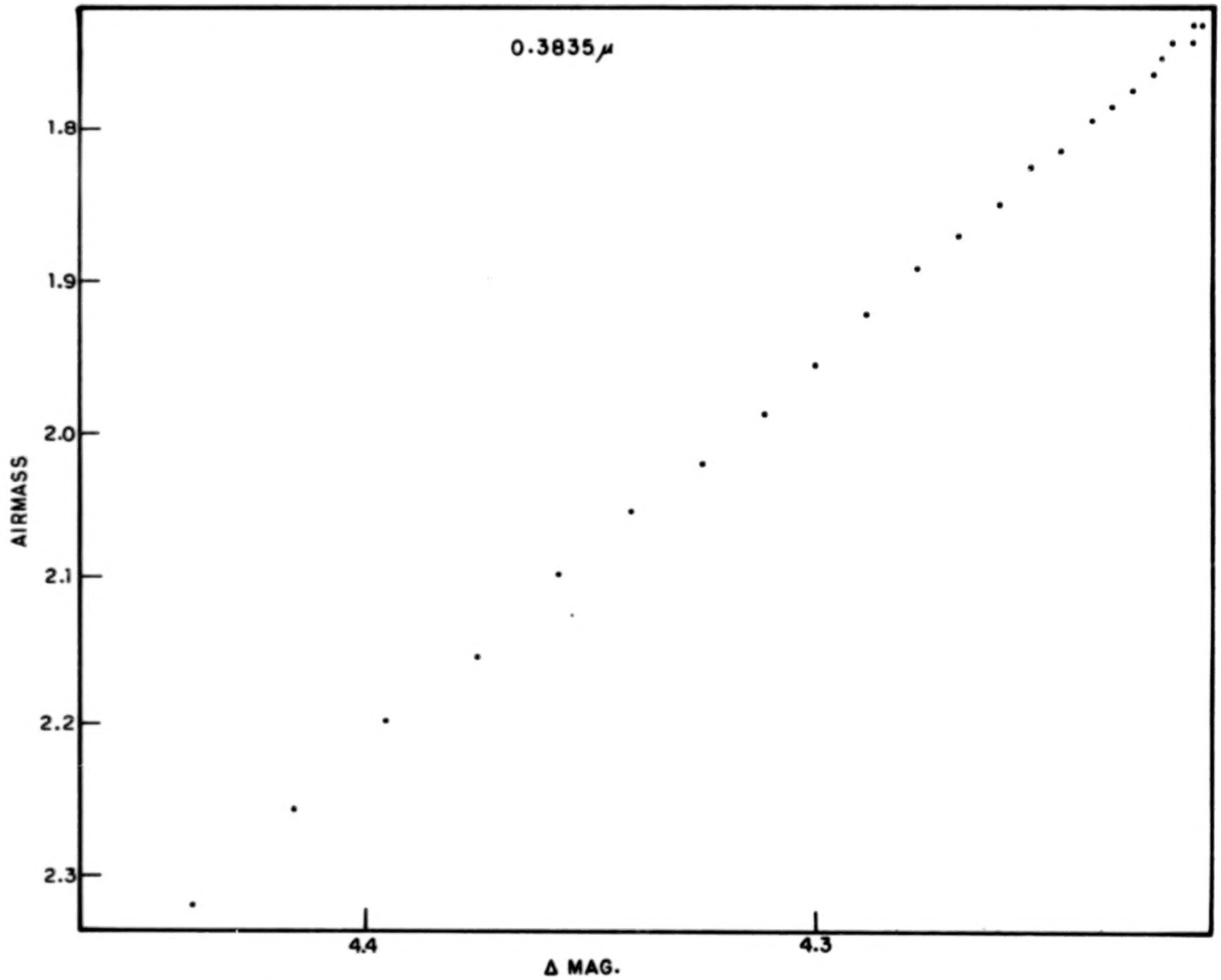


Figure 2. Bouguer curve from solar radiometer at wavelength 0.3835 microns.

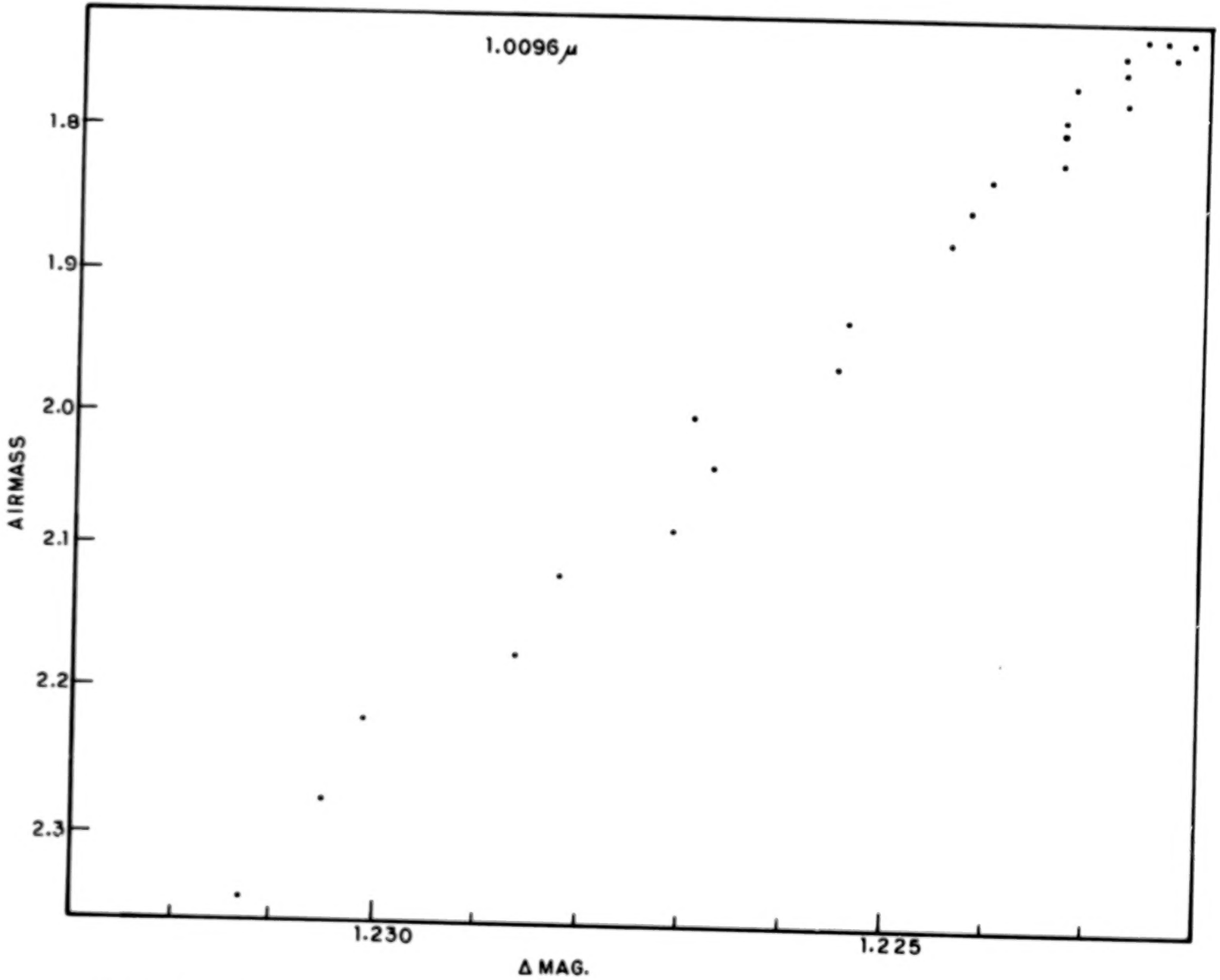


Figure 3. Bouguer curve from solar radiometer at wavelength 1.0096 macrons.

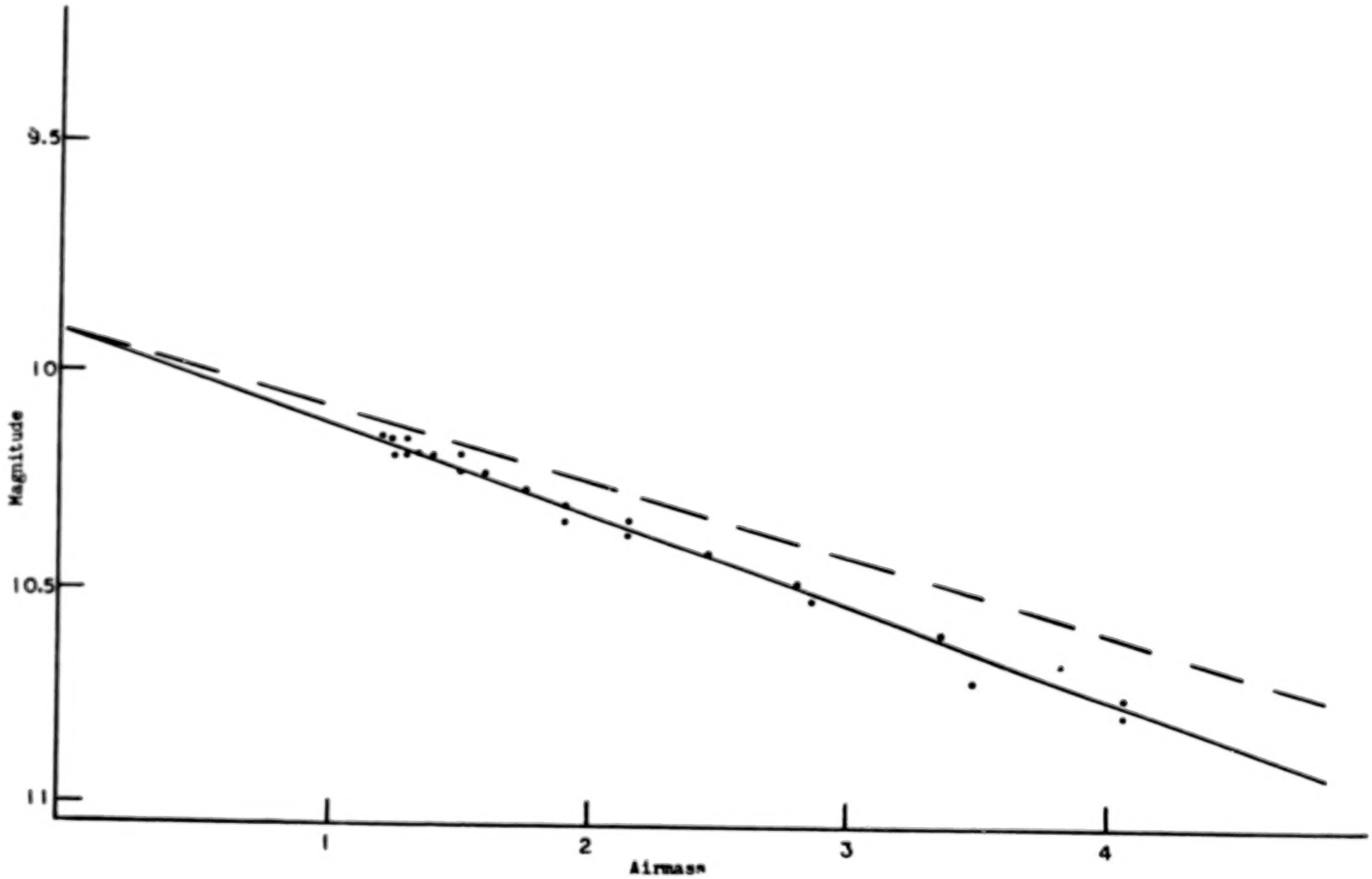


Figure 4. Bouguer curve from night time measurements with Mount Laguna 16-inch telescope at wavelength 0.4605 microns.

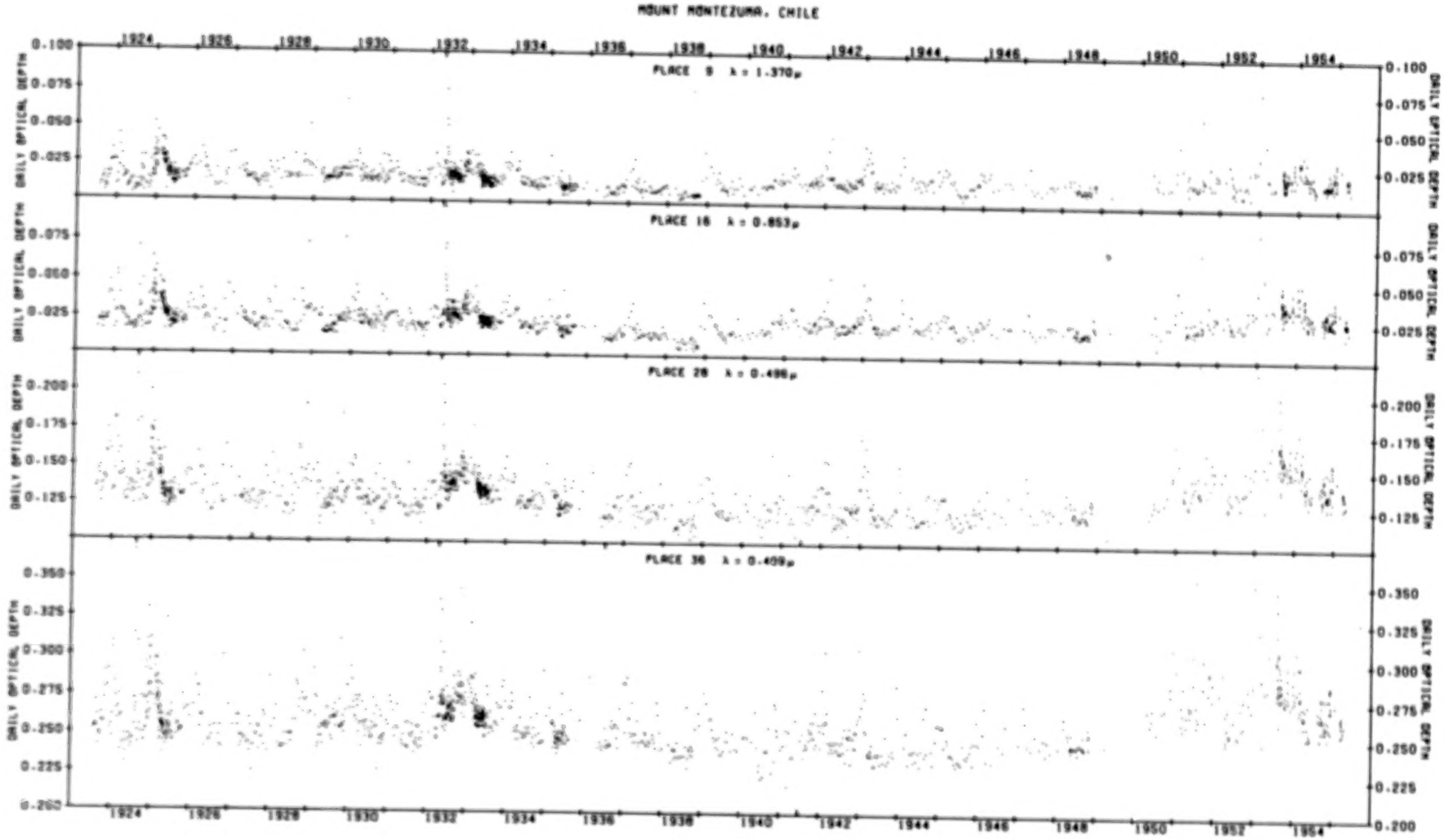


Figure 5. Time series of optical depth measurements at four different wavelengths from Smithsonian data.

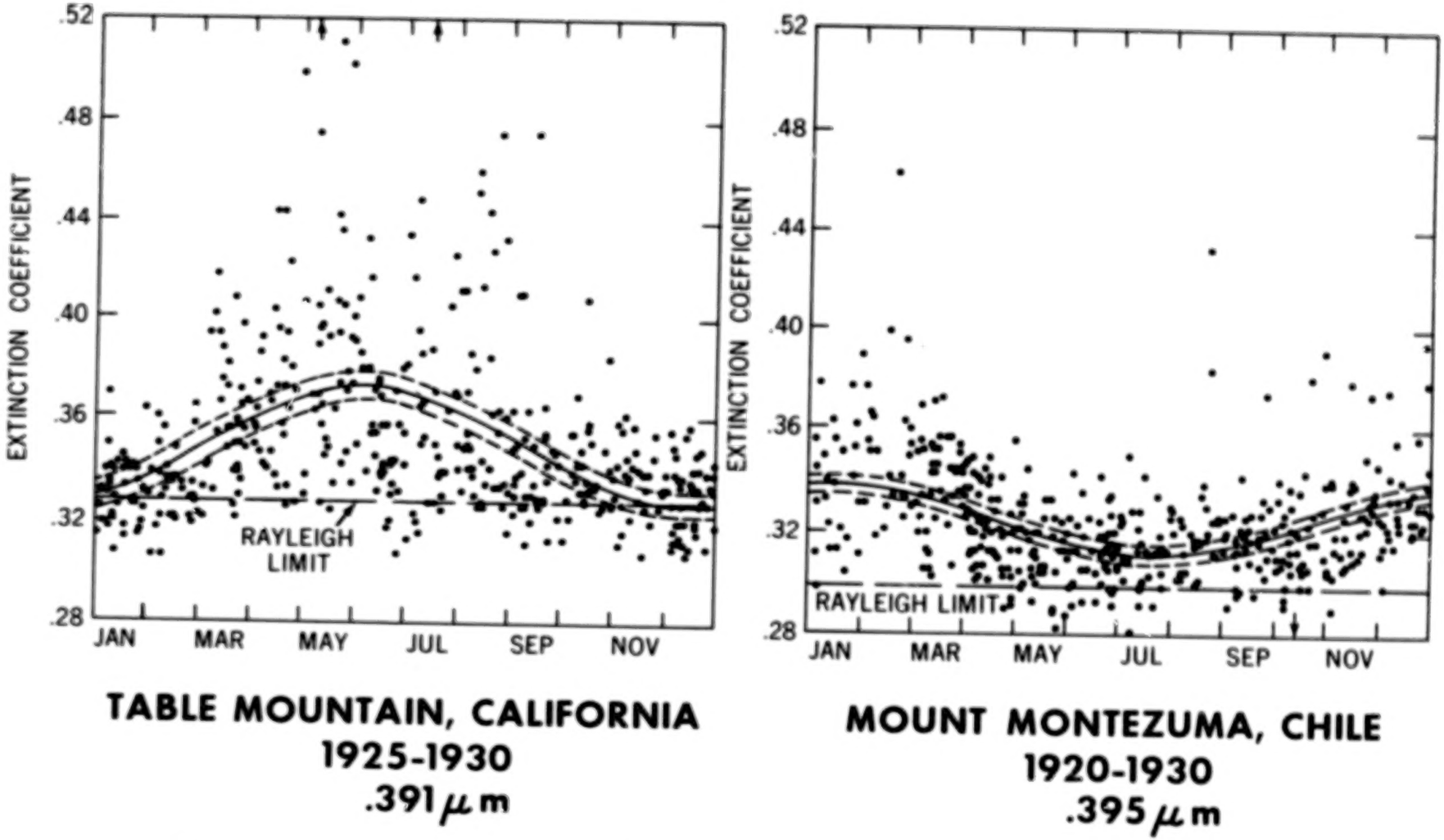


Figure 6. Seasonal variation in extinction coefficient for both a northern and southern hemisphere site illustrated from Smithsonian data.

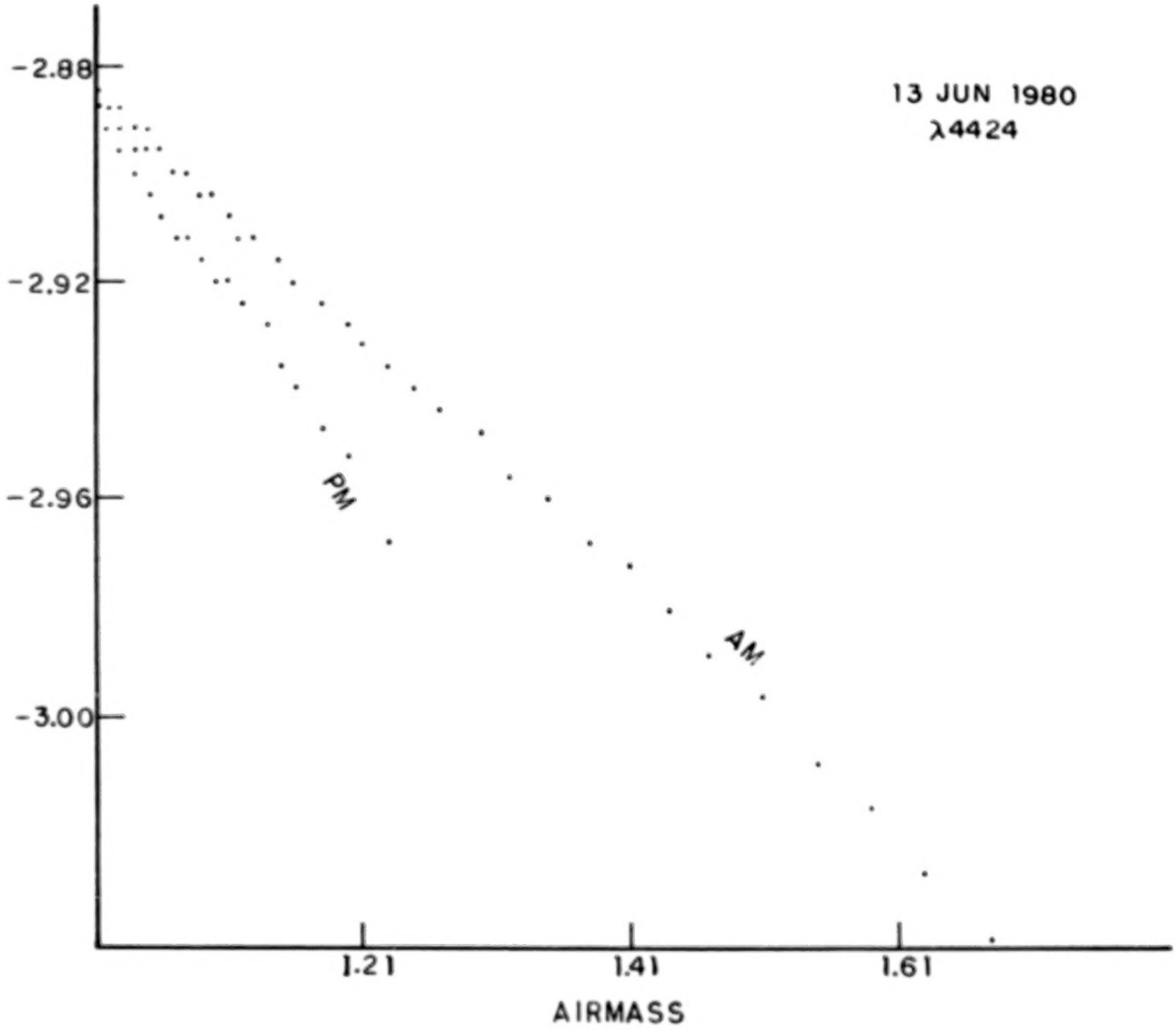


Figure 7. Bouguer curve from solar radiometer illustrating time variable extinction.

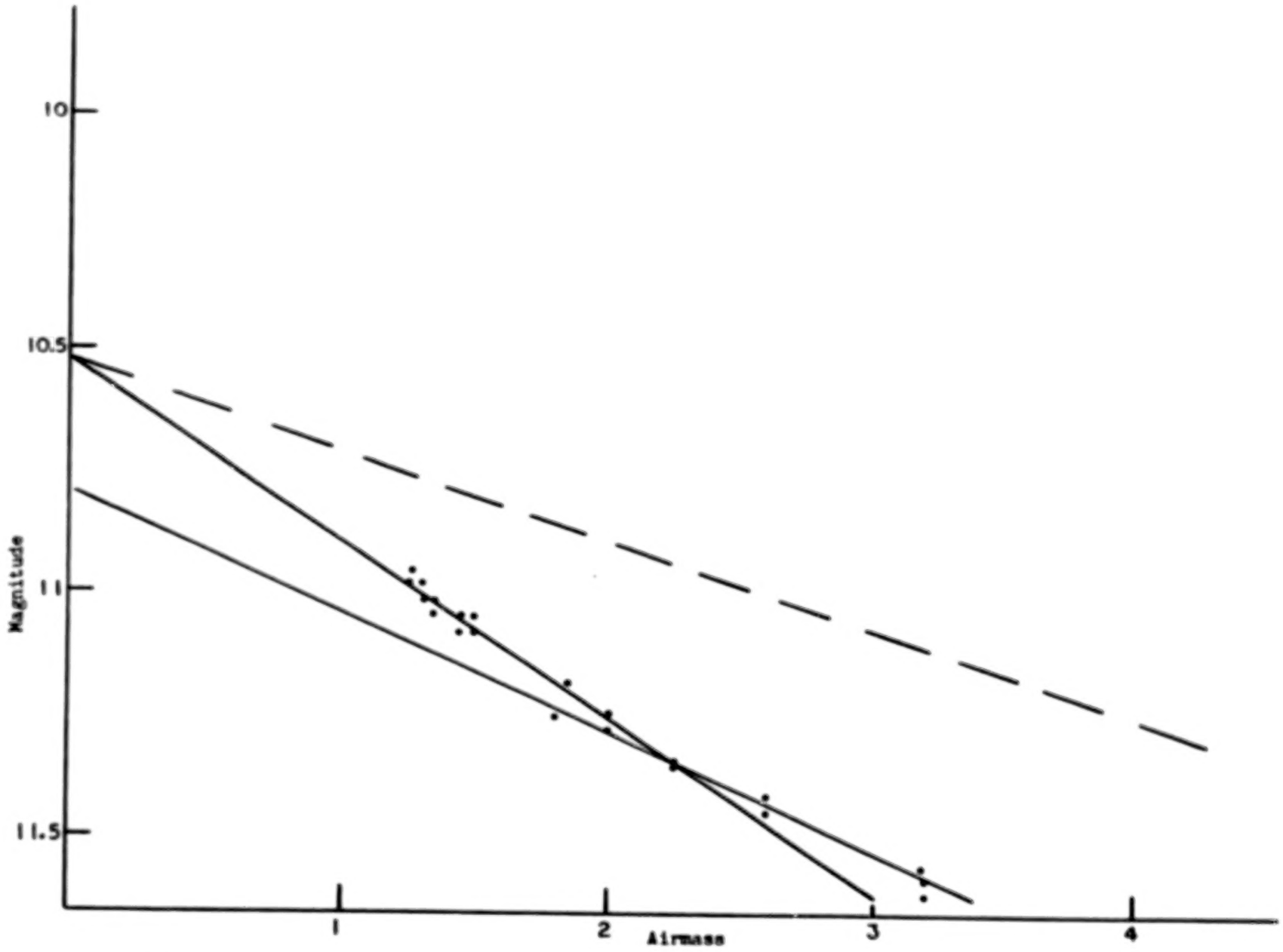


Figure 8. Time variable extinction from night time measurements from Mount Laguna 16-inch telescope.

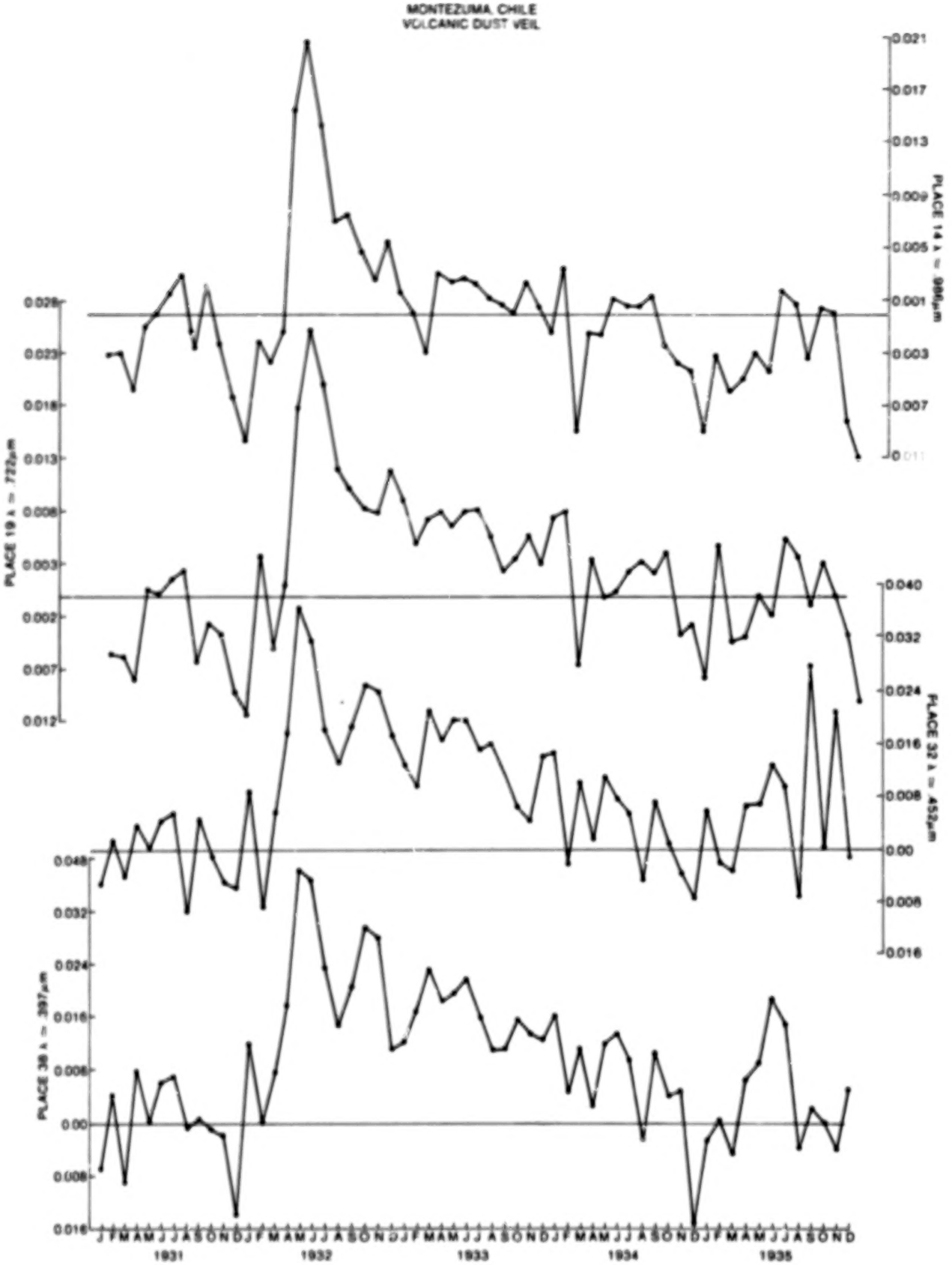


Figure 9. The effect of the eruption of Mount Quizapu on atmospheric extinction shown at four wavelengths.

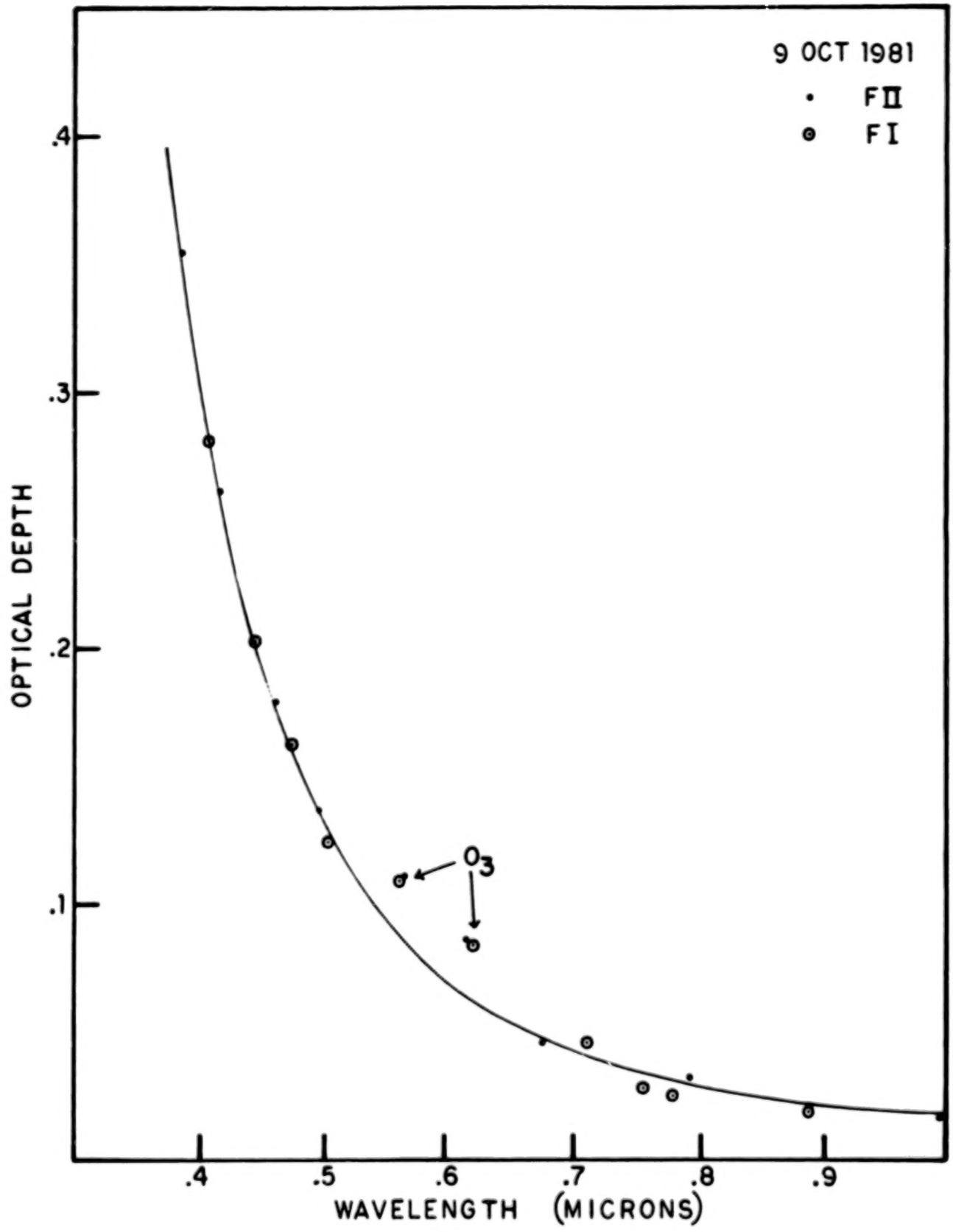


Figure 10. Rayleigh and aerosol modelling of the atmospheric optical depth (solid line); plotted points are observed data.

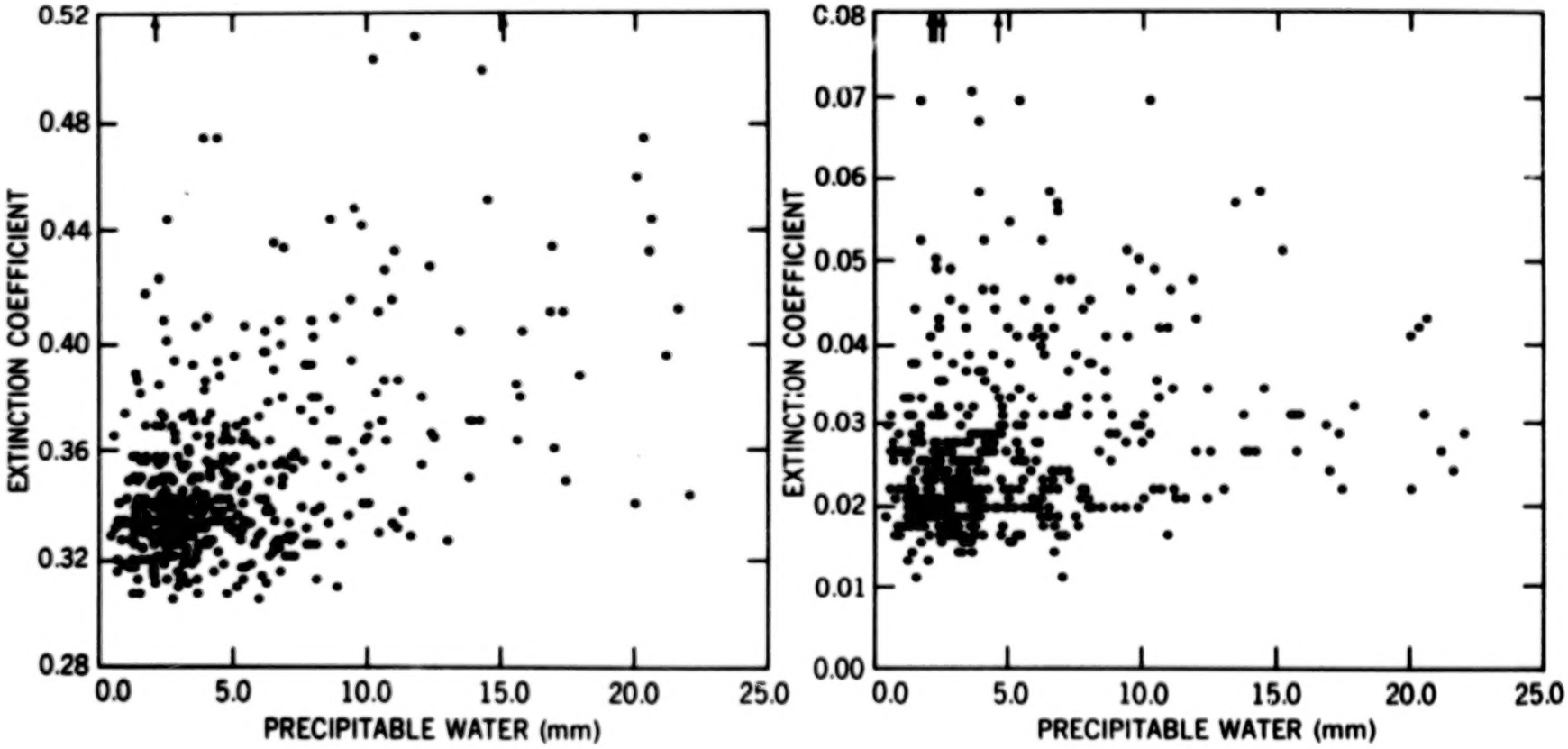


Figure 11. The correlation between extinction and precipitable water vapor from Smithsonian measurements.

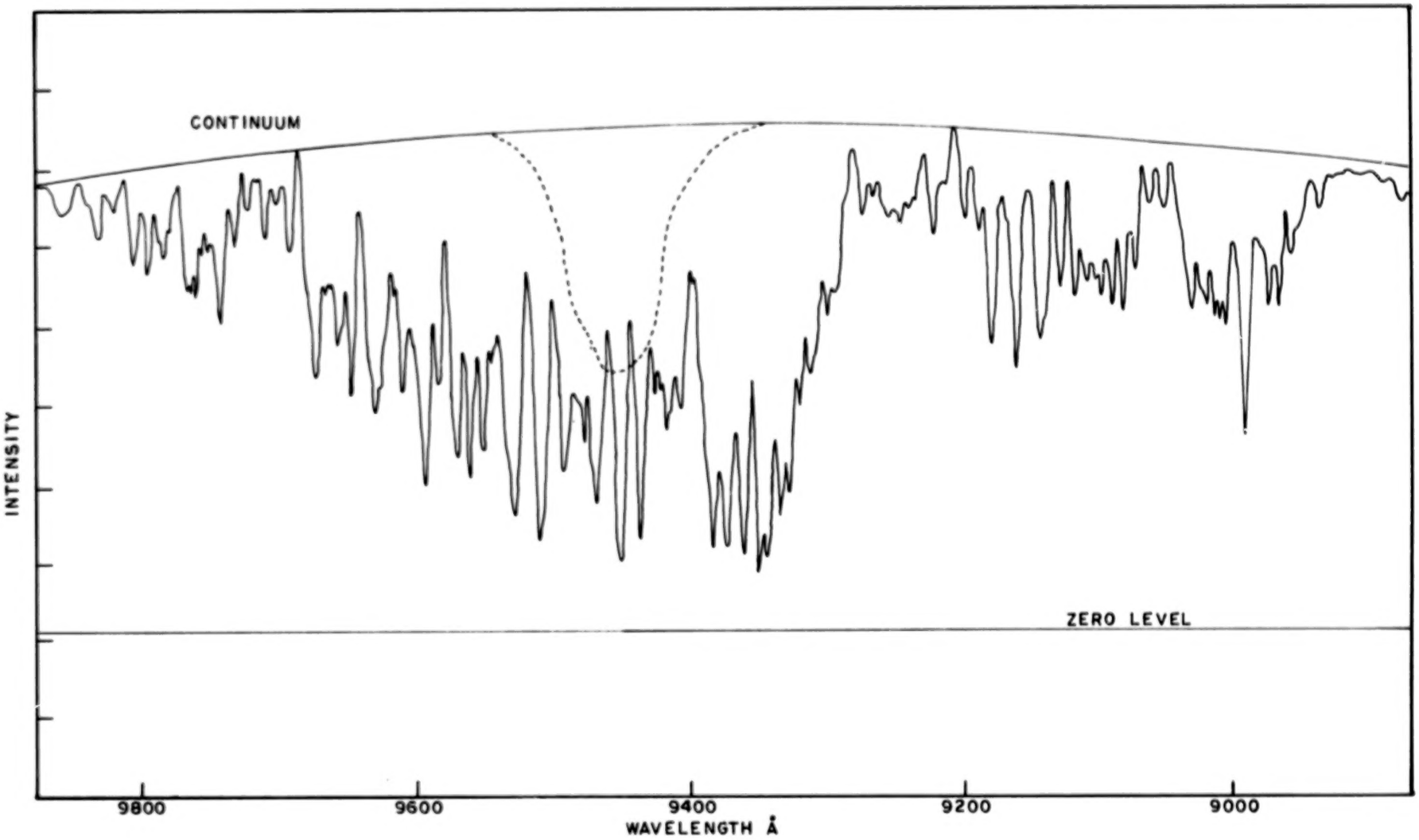


Figure 12. Spectral scan of solar irradiance in the region of the 0.94 micron water vapor bands. Smooth line is the assumed continuum, and the dashed line is the spectral transmission of the filter used to measure water vapor.

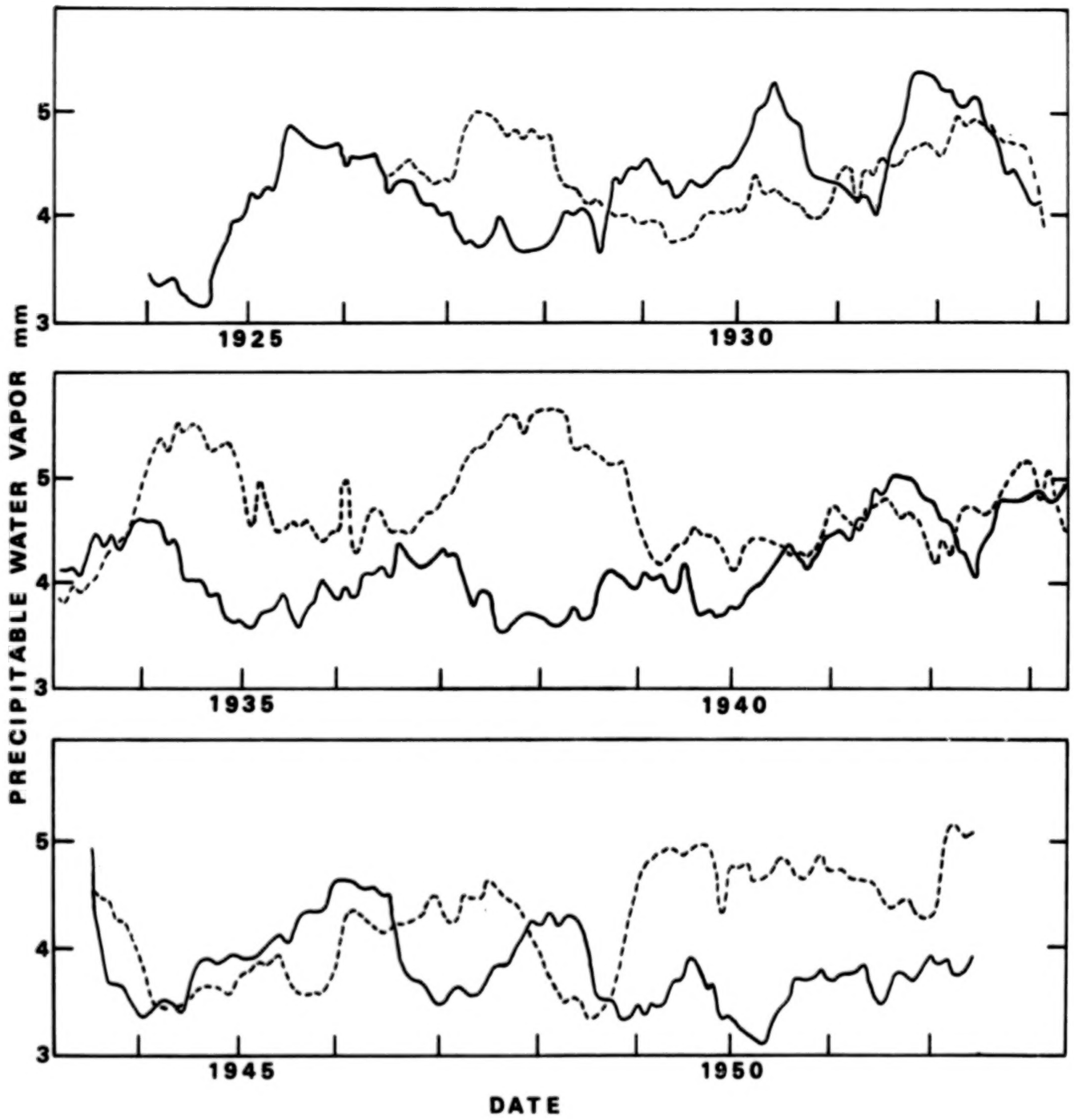


Figure 13. Time series showing the variability of water vapor (seasonal variation has been removed). Solid line represents data from Mt. Montezuma, Chile, and dashed line from Table Mountain, California.

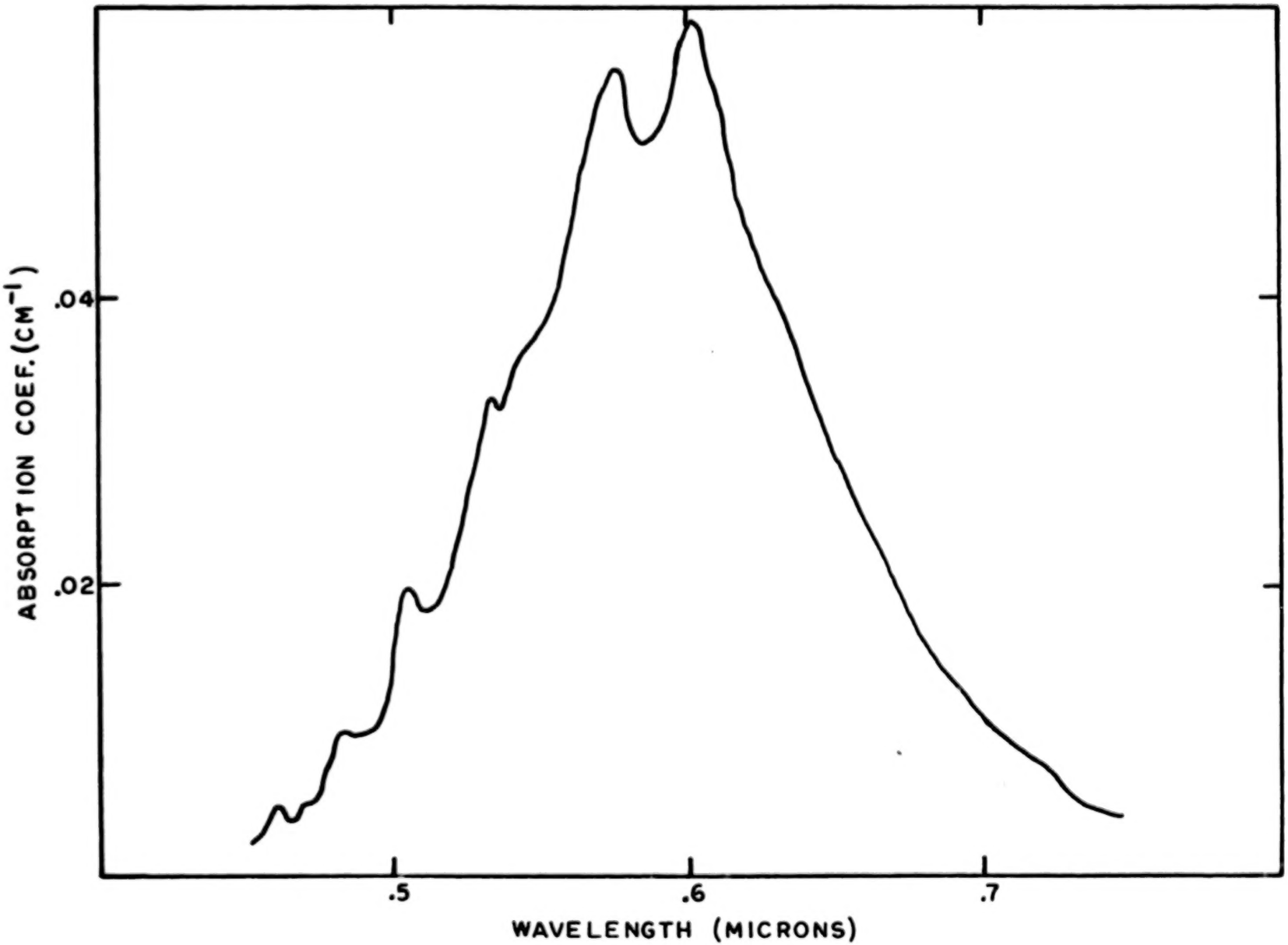


Figure 14. The wavelength extent of the Chappuis ozone band.

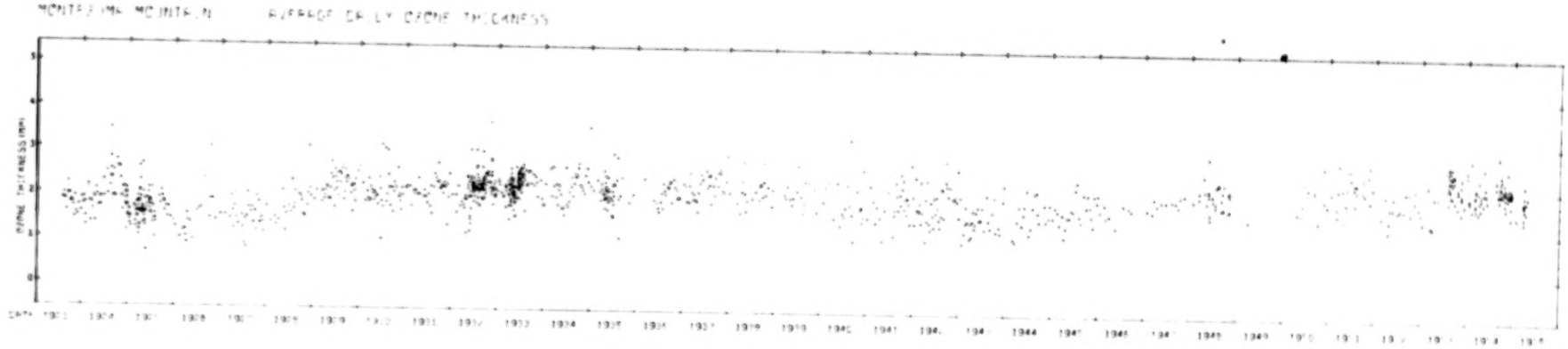


Figure 15. Time series showing variation in atmospheric ozone from Smithsonian data at Mt. Montezuma, Chile.

REDUCTION TO OUTSIDE THE ATMOSPHERE AND STATISTICAL TESTS USED IN GENEVA PHOTOMETRY

F. Rufener
Geneva Observatory

1. INTRODUCTION

Twenty five years ago, the Geneva Observatory undertook the ambitious task of creating a precise photometric system. The beginning was difficult and slow, but we are now proud of having collected a homogeneous sample of photometric data in seven colours for more than 23'000 stars (Rufener, 1981). The experience I have gained during that venture has confirmed several convictions which I have maintained for a long time. The analytical and discriminatory potentials of a photometry obviously result from the localization of the passbands in the spectrum; they do, however, also depend critically on the precision attained. This precision is the result of two different types of precautions.

a) The "hardware"-related precision. In addition to the contribution of the quality and reliability of the device which measures the photometric signal, precision is also conditioned by the exactness with which the passbands are defined and maintained, independently of all possible perturbations. Any negligence at this level creates an overwhelming handicap which affects the subsequent true accuracy and significance of a photometric quantity.

b) The "software"-related precision. It results from the quality of the techniques of acquisition of the measurements, and depends on the models used for their numerical treatment in view of reduction to their values outside the atmosphere and of correlation with an excellent standard of reference.

I do not intend here to convince you that all these aspects have been treated perfectly well in Geneva photometry (this would be presumptuous and certainly not true), nor do I wish to undertake a proselytic action in favour of this photometry which has enabled us to obtain some original results (Golay, 1980). I shall limit my presentation to two procedures which, in my opinion, contribute in an efficient manner to acheiving greater precision. These two procedures, it seems to me, are still rather poorly known in spite of the fact that they have both been in use for a long time in Geneva.

2. REDUCTION TO OUTSIDE THE ATMOSPHERE

The organization of the measurements in view of enabling the correction of atmospheric absorption is an important subject which can greatly contribute to the accuracy of their values reduced to outside the atmosphere. We have given up the simplistic model of the Bouguer line in favour of a somewhat more realistic model. Moreover, the effects due to the displacement of the "effective" wavelength relative to the atmospheric absorption, which are quite appreciable in the case of a wide band filtering the light of stars with very different temperatures, have been corrected in an original manner with a systematic application to all filters.

It is now necessary to discuss a few details before presenting these methods. Let us consider a passband described by the following characteristics (Fig. 1):

λ_o its mean wavelength

$$\lambda_o = \int \lambda \phi(\lambda) d\lambda \; / \; \int \phi(\lambda) d\lambda$$

μ an estimator of its half-width, so that

$$\mu^2 = \int (\lambda - \lambda_o)^2 \, \phi(\lambda) d\lambda \; / \; \int \phi(\lambda) d\lambda$$

Φ the brightness of the passband expressed in magnitudes

$$\Phi = -2.5 \log \int \phi(\lambda) d\lambda$$

By adopting the theoretical approaches proposed by B. Strömgren (1937), I. King (1952) and F. Rufener (1964) we obtain a formulation which describes the apparent magnitude measured at ground-level (m_z) by means of terms which contain the above-mentioned passband characteristics, the properties of the mean extinction of the observatory site and the energy distribution of the source.

$$m_z = \underbrace{m_o(\lambda_o) - 0.543\mu^2 \frac{E''(\lambda_o)}{E(\lambda_o)} + \Phi}_{m_o}$$

$$+k(\lambda_o)F_z \left[1 + \frac{n_{\lambda_o}(n_{\lambda_o}+1)}{2} \left(\frac{\mu}{\lambda_o}\right)^2 - n_{\lambda_o} \frac{\mu^2}{\lambda_o} \frac{E'(\lambda_o)}{E(\lambda_o)} \right] \qquad (1)$$

$$-0.46 \, n_{\lambda_o}^2 \left(\frac{\mu}{\lambda_o}\right)^2 k^2(\lambda_o) F_z^2$$

Formula (1) retains the principal terms of a series development; it shows how the width of the passband (μ) intervenes and complicates the classical Bouguer formula. Hereafter we give the definitions of each of the factors involved:

m_z: The heterochromatic magnitude measured at ground level. This is the magnitude for the passband considered which depends on the crossed air mass F_z along the line of sight.

m_o: The same magnitude reduced to outside the atmosphere.

$m_o(\lambda_o)$: The monochromatic magnitude at the mean wavelength λ_o, reduced to outside the atmosphere.

$E(\lambda_o)$: The spectral energy distribution of the observed star and its first ($E'(\lambda)$) and second ($E''(\lambda)$) derivatives. The values at the mean wavelength λ_o are considered. We verify the following relation for the ratio $E'(\alpha_o)/E(\lambda_o)$:

$$\frac{E'(\lambda_o)}{E(\lambda_o)} = \frac{1}{1.086\lambda_o^2}\left[\frac{dm(\lambda)}{d\,1/\lambda}\right]_{\lambda_o} = \frac{1}{1.086\lambda_o^2}\,\frac{C^o_{1-2} - \phi_1+\phi_2}{1/\lambda_1 - 1/\lambda_2}$$

where $m(\lambda) = -2.5 \log E(\lambda)$; $\lambda_1 \leqslant \lambda_o \leqslant \lambda_2$

and C^o_{1-2} is a colour index with $C^o_{1-2} = m(\lambda_1) - m(\lambda_2)$

$k(\lambda_o)$: The value at λ_o of the atmospheric extinction law of the site.

n_{λ_o} : The value at λ_o of the ratio $-\frac{\Delta \log k(\lambda)}{\Delta \log \lambda}$

(If n is constant, we have $k(\lambda) = a\lambda^{-n}$).

F_z : The air mass along the line of sight during the observation. It is defined as the number of times one has, along the line of sight, the quantity of air seen in the direction of the zenith.

We see that a good knowledge of the passbands of the photometer, of the mean extinction at the observatory site and of the energy distribution of the star observed (i.e. of an appropriate colour index) allows a numerical estimate of the factors in formula (1). We can therefore calculate numerically for each filter the terms α, β and γ which are used here to simplify the expression of formula (1).

$$m_z = m_o + k(\lambda_o)F_z\,[1+\alpha_{\lambda_o} + \beta_{\lambda_o} C^o_{1-2}] + \gamma_{\lambda_o} F_z^2 \qquad (2)$$

This relation replaces the Bouguer line. If we consider the passband to have a negligible width we will have $\mu = o$ and, of course, $\alpha = \beta = \gamma = o$. Relation (2) is then reduced to the classical Bouguer line:

$$m_z = m_o + k(\lambda_o)F_z \qquad (3)$$

It is customary to use the Bouguer line (3) to determine the atmospheric extinction $k(\lambda_o)$ corresponding to the observations in the passband λ_o. During the night, several measurements of the ground-level magnitude (m_{z_i}) of the star are made while its air mass changes significantly (F_{z_i}). The application of a linear regression optimized by least squares allows the extinction value $k(\lambda_o)$ and the magnitude of the star outside the atmosphere (m_o) to be computed. This method implicitly relies on three hypotheses: the stability of the photometer, the stability of the star and the stability of the absorption properties together with the isotropy of the atmosphere during the whole night (at least 5 to 7 hours!). According to our experience, the technical precautions described above, together with the choice of a "good" star, enable the first two of these hypotheses to be satisfied. On the other hand, it is rare that the third one is also satisfied.

To illustrate the inconveniences which can result from a slowly variable extinction we present, in Fig. 2, an example of what can happen when the atmospheric extinction decreases during the time necessary to observe two extinction stars: the one ascending (M), the other descending (D).

We assume that the extinction, if not constant, varies slowly and isotropically in the whole solid angle considered. This hypothesis is less demanding than that of Bouguer and is reasonable when the sky is clear and the variations of extinction result mainly from drifts of atmospheric pressure, from the slow evolution of the vertical distribution of the aerosols and from the nocturnal evolution of the states of equilibrium between absorption and emission in the atomic and molecular bands of the night sky. For the demonstration in Fig. 2 we have chosen a slowly decreasing extinction and fixed a priori the $k(\lambda_o, t_i)$ values at the times t_i, distributed over the whole night. At these given times, we simulate consecutive (quasi-simultaneous) observations of each star M and D (m_{Mzi} and m_{Dzi}). Fig. 2 reveals what is observed at ground level for the stars M and D whose magnitudes outside the atmosphere (m_{Mo} and m_{Do}) are assumed to be constant. If these simultaneous observations were to be treated by Bouguer's method, we would obtain two different estimates of the extinction coefficient, both wrong, with the one smaller and the other larger than the true extinction values.

By adopting the hypothesis of a slowly varying isotropic extinction, formula (2) becomes time-dependent with extinction taking the form $k(\lambda_o, t)$. We assume that the terms α, β and γ remain constant and are estimated by a mean extinction ($\bar{k}(\lambda)$). We use relation (2) in the following form:

$$m'_z = m_o + k(\lambda_o,t)F'_z \qquad (4)$$

with

$$m'_z = m_z - \gamma_{\lambda_o} F_z^2$$

$$F'_z = F_z(1+\alpha_{\lambda_o} + \beta_{\lambda_o} C^o_{1-2})$$

Each observation is thus corrected for the band width effects. The "M and D" method proposed by Rufener (1964) consists in obtaining several (n) pairs of observations of two stars, the one called M (ascending) and the other D (descending). We then have two series:

$$\begin{aligned} m'_{Mzi} &= m_{Mo} + k(\lambda_o,t_i)F'_{Mzi} \\ m'_{Dzi} &= m_{Do} + k(\lambda_o,t_i)F'_{Dzi} \end{aligned} \qquad (5)$$

where i = 1, ... n

Following the hypothesis that the atmospheric extinction $k(\lambda_o,t_i)$ is the same for the measurements of the stars M and D made at the time t_i, we can eliminate $k(\lambda_o,t_i)$ and treat the remaining n relations by least squares.

$$\frac{m'_{Mzi} - m_{Mo}}{F'_{Mzi}} = \frac{m'_{Dzi} - m_{Do}}{F'_{Dzi}} \qquad (6)$$

When n>2 (usually 4 or 5), the system (6) determines by least squares m_{Mo} and m_{Do} which are the magnitudes outside the atmosphere of the stars M and D. We then compute the individual $k(\lambda_o,t_i)$ values which are interpolated for the reduction of the other observations of the night to outside the atmosphere.

This M and D method enables the atmospheric extinction to be measured even if it varies slowly during the night. The only assumption is that when a variation takes place, it occurs isotropically, so that the whole solid angle in which the observations are made is affected by the variation of transparency. The effects of Fig. 2 are avoided by the M and D method which allows, under such circumstances, a true determination of m_{Mo} and m_{Do} whereas an application of the Bouguer method to the same data would lead to false results. This M and D method has been applied at the Geneva Observatory for more than twenty years whenever extinction measurements were made. The values obtained are then interpolated for the reduction outside the atmosphere of our program stars, observed through any air mass. When extinction is not measured, we adopt a mean value. We then conduct our observations at constant air mass ($\bar{F}_z$). Relation (2) becomes, in this case

$$m_z = m_o + (\bar{k}(\lambda_o)+\Delta k)\bar{F}_z\,[1+\alpha_{\lambda_o} + \beta_{\lambda_o} C^o_{1-2}] + \gamma_{\lambda_o} \bar{F}^2_z \qquad (7)$$

The unknown deviation (Δk) from the mean extinction ($\bar{k}(\lambda_o)$) which characterizes the observational conditions of the night occurs only as a multiple of $\bar{F}_z$, which is the air mass chosen for that night. The deviations $\Delta k\bar{F}_z$ during the night can then be estimated with the aid of a sufficiently large sample of standard stars. Even if Δk varies a little, it is easy to compensate this variation by smoothing over the Δk values and thus obtain precise magnitudes.

3. RECORDING OF THE PHOTOMETRIC SIGNAL AND THE THREE STATISTICAL TESTS Q, R and G

Computer-controlled photometers with pulse counting have replaced earlier DC amplifier and chart-recorder output photometers. The modern system presents many advantages but lacks the visual control provided by the chart-recorder, which also gives a chronological record of the measurement. The following three statistical tests can easily be calculated in real time and, recorded with the data, they provide good indications about any slow or rapid anomaly in the signal. They enable us to discriminate between the photometric signal and the atmospheric scintillation, slow changes or oscillations (due to variable absorption, decentering etc.) and asymmetry (spikes, drops). The only requirement is that the total integration time be broken into many (> 100) sub-intervals, hopefully without dead time. These tests were devised while building the two-channel photometer for the Geneva Seven-Colour Photometric System (Burnet and Rufener, 1979) and have been in continuous use since 1977. They have been of considerable help in eliminating data affected by the many pitfalls of photometrical observations, often long before they could be detected otherwise. The following description is taken from a paper by Bartholdi et al. (1984).

Let the total integration time T be broken into n equal sub-intervals ΔT_i such that $T = \Sigma\,\Delta T_i = n\Delta T_i$, N_i being the count during ΔT_i. We can then establish the following variables for each integration:

$$N = \Sigma\, N_i/n$$

$$\sigma^2 = \Sigma\,(N_i-N)^2/n$$

$$Q = \sigma^2/N \qquad (8)$$

$$R = \Sigma\,(N_i-N_{i-1})^2/(2n\sigma^2) \qquad (9)$$

$$G = \Sigma\,(N_i-N)^3/(n\sigma^3)-1/N^{1/2} \qquad (10)$$

In practice, we frequently use the counting rate $N' = Nn/T$.

For on-line computation, it is easier to replace the preceding formulas by those obtained with the following, continuously updated sums:

$$A = \Sigma N_i$$
$$B = \Sigma N_i^2$$
$$C = \Sigma N_i^3$$
$$D = \Sigma(N_i - N_{i-1})^2$$

Then

$$N = A/n$$
$$\sigma^2 = (nB-A^2)/n^2$$
$$Q = (nB-A^2)/nA$$
$$R = 0.5\ nD/(nB-A^2)$$
$$G = (n^2C-3nAB+2A^3)/(nB-A^2)^{3/2}n^{3/2}-(n/A)^{1/2}$$

These two sets of formulas are strictly equivalent. In the second, only the four running sums (A to D) should be updated after each sub-integration. None of the subresults (N_i) needs to be stored.

4. TEST SIGNIFICANCE

a) Atmospheric scintillation

Q is the ratio of the observed noise to the theoretical photon (poissonian) noise. Q should be greater or equal to 1. Q greater than 1 implies some extra noise, principally due to atmospheric scintillation. Let us assume that the photon flux is modulated by a factor $1+s(t)$ where $s(t)$ is a stationary random process of mean 0 and variance $v^2 > 0$. $s(t)$ is independent of the poissonian fluctuations. $N_{i\,0}$ is the poissonian sample, with only photon noise, having N as mean and variance. Then the observed counts $N_i = N_{i\,0}(1+s_i)$, where s_i is the integration of $s(t)$ over ΔT_i. Introducing this relation into (1), we find that

$$Q = 1+s^2\ N$$

where

$$s^2 = \left\langle \left[\int_{\Delta T_i} s(t)dt\right]^2 \right\rangle = \langle s_i^2 \rangle \qquad (11)$$

Figure 3 shows a typical sample of Q values as a function of N, when measured simultaneously during 64 seconds, confirming the validity of (11).

From Q and N we can deduce a fairly good estimate of s^2 which we assume to be representative of the atmospheric scintillation:

$$s^2 = (Q-1)/N \qquad (12)$$

Notice that if $s^2 N \ll 1$, Q will not differ from 1 and s^2 will not be significant. That is to say that the photon noise will largely dominate the scintillation. On the other hand, if $s^2 N > 1$, the scintillation dominates. s^2 is a function of time, air-mass and telescope aperture.

Using Reiger's theory, Young (1974) has shown how s^2 varies with air-mass. Figure 4 presents two examples confirming that this variability satisfies the relation $s^2 = s_0^2 F^k$ with $k \simeq 3$. The same author has discussed the telescope aperture dependency of s^2. We derive a similar power law from figure 5. The exponent of -1.7 is in-between Young's theoretical values. When very bright stars are observed with small telescopes, it is confirmed that the integration time has to be very large if high precision is desired. s^2 also depends directly on $\Delta T_i \simeq 15$ ms.

As only observed quantities appear in the numerator and the denominator of R and G, both will be essentially insensitive to scintillation as long as it is symmetric around the mean and uncorrelated from one sub-interval to the next.

A linear trend modifies Q exactly as an atmospheric scintillation and is not distinguishable from a true scintillation when using Q alone. Notice also that, as for the scintillation, a linear trend will affect Q only in case of large N, except if this trend is very strong.

Finally, we can calculate the effect on Q of a single spike (or drop) of v counts occurring during a single sub-interval.

$$Q' \simeq (QN+v^2)/(nN+v) \qquad (13)$$

b) Slow changes or oscillations

The second test was used a long time ago by Abbe (1906) to detect trends in statistical series. More recently it was extensively studied by von Neumann (1941) and generalised by Allan (1966). If the N_i are randomly uncorrelated, a fundamental assumption for the Poissonian process (Haight, 1967), R should be equal to 1. If the data show slow trends, R will be smaller than one; conversely, R will be greater than one if the signal oscillates. In all cases $0 < R < 2$.

Contrary to the test Q which mixes scintillation and trend effects, the R test decreases with a trend and stays quite insensitive to atmospheric scintillation. These properties are well shown in Fig. 8 and 12. Notice that R is very insensitive to single spikes or drops, as their behaviour is identical in the numerator and the denominator of (9).

c) Asymmetrical spikes and loss in the signal

The last test is simply the normalised third moment of the distribution, also called the skewness. The second term in the G-test (10) is the theoretical value for a Poisson distribution. It is essential for small N. If all expected positive and negative deviations are compensated, then G will be zero. It will be positive if spikes are present and negative for drops.

Suppose one of the sub-intervals has received v (positive or negative) extra counts, while all others are normal, and that there is no scintillation ($N > s^2 N^2$);

then

$$G \cong v^3/n(N+v^2/n)^{3/2} \qquad (14)$$

and v can thus be estimated with:

$$v = (N+v^2/n)^{1/2}(nG)^{1/3} \qquad (15)$$

Both (13) and this formula can be used for computing a fairly good estimate of even a small spike above the normal photon noise. Note that they contain implicitly v in both formulas, which consequently have to be iterated, starting with v = 0. The simultaneous application of (13) and (15) to weak signals leads to the detection of rare events as shown in Fig. 6.

While the first two tests become more sensitive with increasing N, G will be particularly good for detecting small spikes or drops in weak signals. This is well displayed in Fig. 13. Notice also that a linear trend, if it is symmetric above and below the mean, has no effect on G. Of course, any higher-order, non-symmetric trend will be detected immediately.

5. CONFIDENCE INTERVAL AND EXPERIMENTAL RESULTS

In the paper by Bartholdi et al. (1984) we have proposed to build a confidence interval around Q, R and G such that if the observed value falls outside this interval, a warning message can be given to the observer and eventually the integration restarted or rejected.

The two-channel Geneva Seven-Colour Photometer described by Burnet and Rufener (1979) is equipped with a fast rotating filter wheel. The integration time is generally fixed at 3.8 seconds, divided into n = 256 sub-intervals which correspond to each pass of a given filter for a given channel. Under these conditions, a complete set of integrations (seven filters, two channels) takes 64 seconds, including dead times. This main integration will be repeated automatically as long as the total number of counts for each filter is not sufficient. Each integration of 64 seconds is recorded on magnetic tape, along with times and the three statistical tests for each of the seven filters and both channels. Thus, any integration of 64 seconds, or any one of the 14 subchannel integrations can be dropped if some anomaly is found in- or off-line without losing the rest of the measurement.

For illustration, the distribution of our three tests taken from 35 nights in spring 1982 is presented in Figs. 7, 8 and 9. The variation of the scintillation during a single night in February 1978 is shown in Fig. 10. It should be noted that none of these documents results from a specific observing programme. In fact, these nights were entirely devoted to very faint stars or to the monitoring of small amplitude variables.

REFERENCES

Abbe, E.: 1906, Werke, Bd. 2, Iena 55

Allan, D.W.: 1966, Proc. IEEE 54, 221

Bartholdi, P., Burnet, M., Rufener, F.: 1984, Astron. Astrophys. (in press)

Burnet, M., Rufener, F.: 1979, Astron. Astrophys. 74, 54

Golay, M.: 1980, Vistas in Astronomy 24, 141

Haight, A.F.: 1967, Handbook of the Poisson Distribution, John Wiley and Sons Inc., New York

King, I.: 1952, Astron. J. 57, 253

Rufener, F.: 1964, Publ. Obs. Genève, Ser. A, 66

Rufener, F.: 1981, Astron. Astrophys. Suppl. 45, 207

Strömgren, B.: 1937, in "Handbuch der Experimentalphysik" 26, 392

von Neumann, J., Kent, R.H., Hart, B.I.: 1941, Annals of Math. Stat. 12, 153

von Neumann, J.: 1941, Annals of Math. Stat. 12, 367

Young, A.T.: 1974, in Methods of Experimental Physics 12, part A, ch. 3, N. Carlton Ed. Academic Press, New York

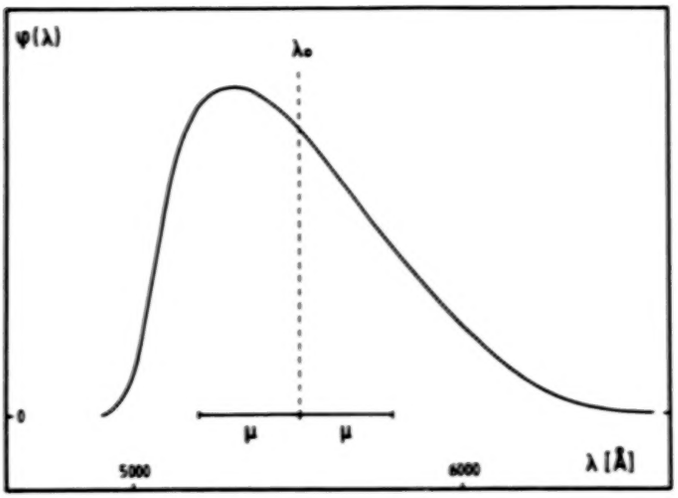


Fig. 1

The spectral response of a typical passband (the V filter of Geneva photometry) with its mean wavelength λ_o and the estimator μ of its half-width.

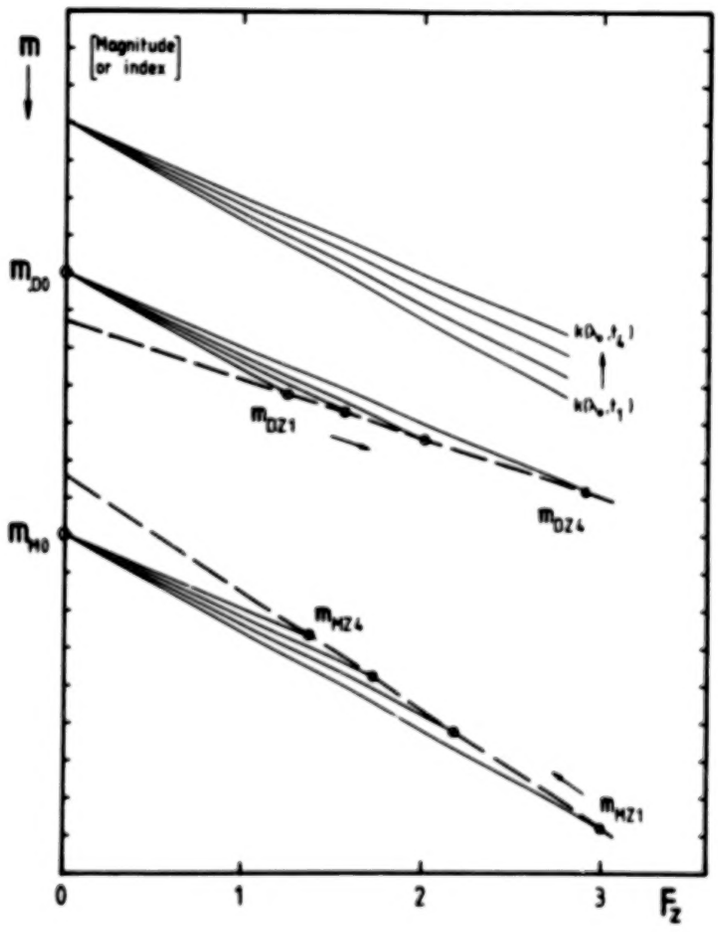


Fig. 2

Simulation of a night with variable extinction. The adopted decreasing values (0.260, 0.240, 0.220, 0.200) of the instantaneous extinction coefficient are shown at the top of the figure. The m_{Mo} and m_{Do} correspond to the true values outside the atmosphere of the stars M and D (these can be unknown). The real observations made at the times t_i would be close to the synthetic values m_{Mzi} and m_{Dzi} represented here by open circles. The two Bouguer lines which one would be tempted to adopt are represented by broken lines. The M+D method described in the text allows to determine the values m_{Mo} and m_{Do} and the i values $k(\lambda, t_i)$. The tick marks of the ordinate scale correspond to 0.1 mag.

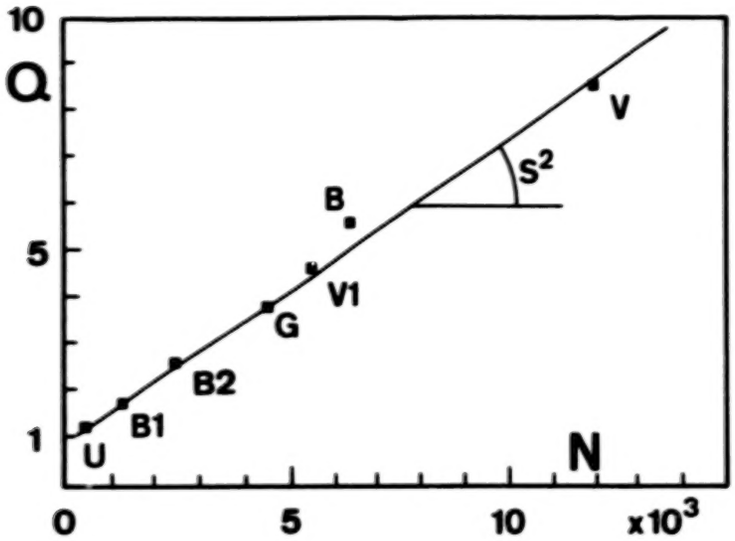


Fig. 3

We have plotted the value of Q versus N, measured simultaneously and corresponding to the seven filters of the Geneva Photometric System. It shows how well the model fits and its independence on colour when the telescope aperture is .7 m. In this example s^2 = 6.4 10^{-4}

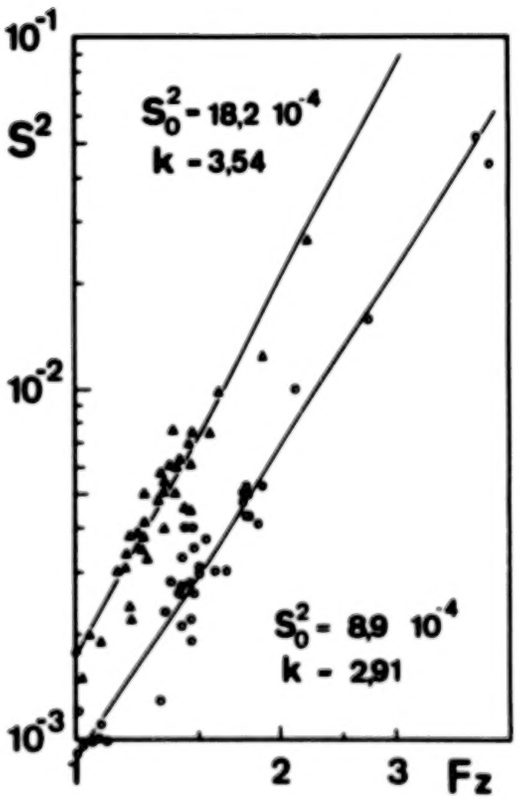


Fig. 4

This figure shows how the atmospheric scintillation varied with air-mass during two different good nights. The lowest set corresponds to a quiet night. The straight line represents a linear fit, assuming that the scintillation varies as $s^2 = s_0^2 F_z^k$. Observations were made with a .4 m telescope.

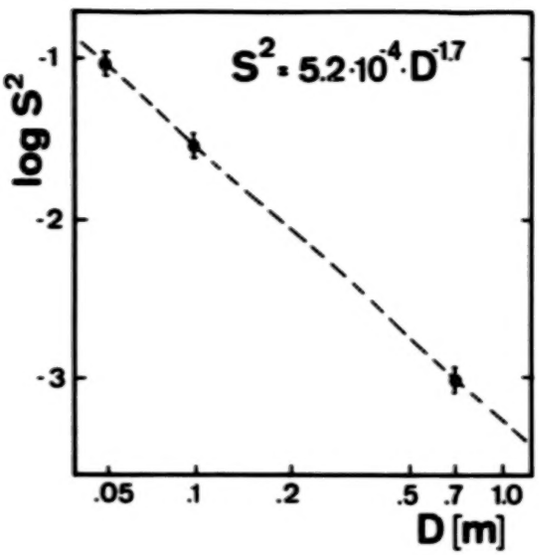


Fig. 5

Atmospheric scintillation as a function of telescope aperture. During three nights we stopped the telescope down to .1 and .05 m with diaphragms exchangeable in a short time. The three points fall on a power line as expected.

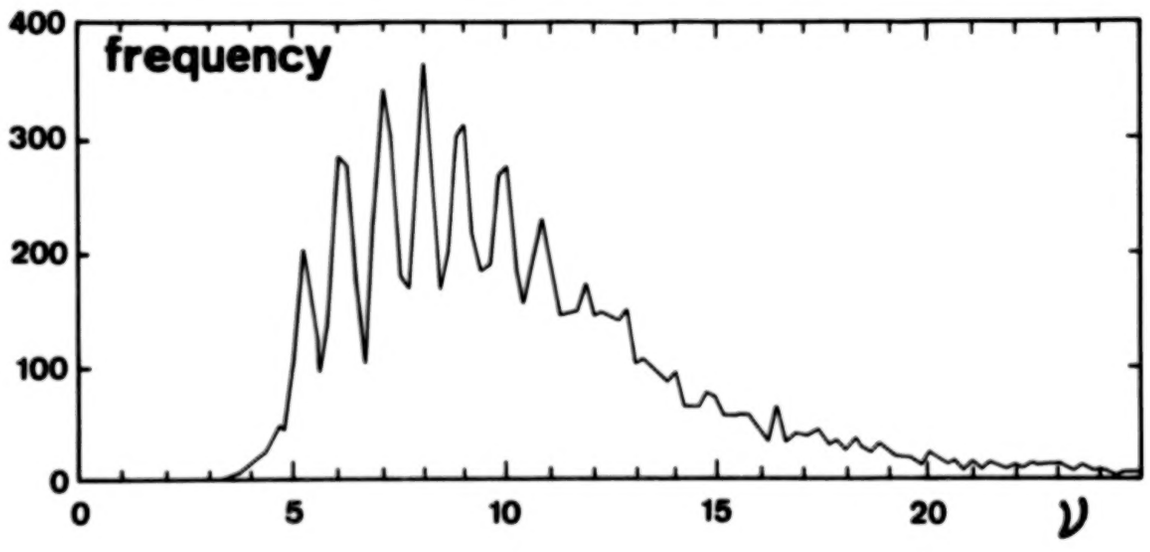


Fig. 6

Frequency distribution of the spikes recovered, using iteratively the formulas (13) and (15). v was calculated with one decimal figure. The clear quantization of v for integer values confirms the ability of these expressions (only integrations with N' < 30 cps were considered).

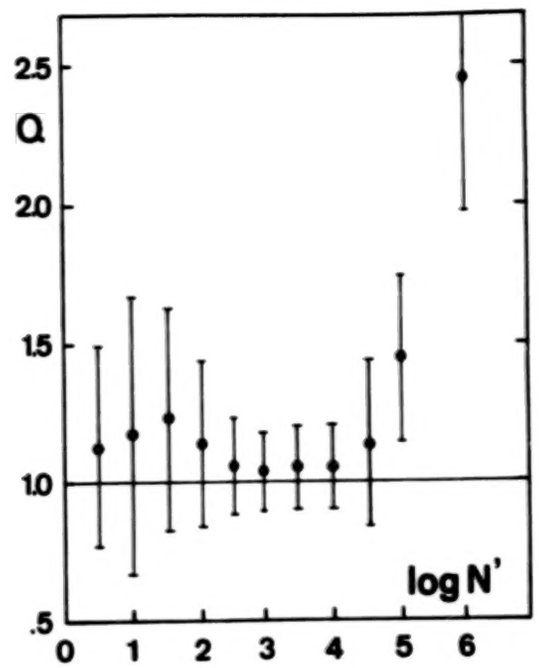


Fig. 7

Mean value of observed Q as a function of log N'. The experimental conditions were: n = 256 samples, ΔT_i = .015 s, telescope aperture = .7 m. 20.000 integrations for each channel and filter recorded during 35 nights are involved. This plot can be divided into three parts:

1. For counting rate N'<200 cps, small perturbations distorted the plot.

2. For 200<N'<20.000 the plot is flat as expected. The observed and theoretical standard deviations agree well.

3. For N'>20.000 cps, the atmosphere scintillation becomes dominant over the photon noise. This part of the plot corresponds to a mean scintillation of $s^2 = 8.10^{-4}$

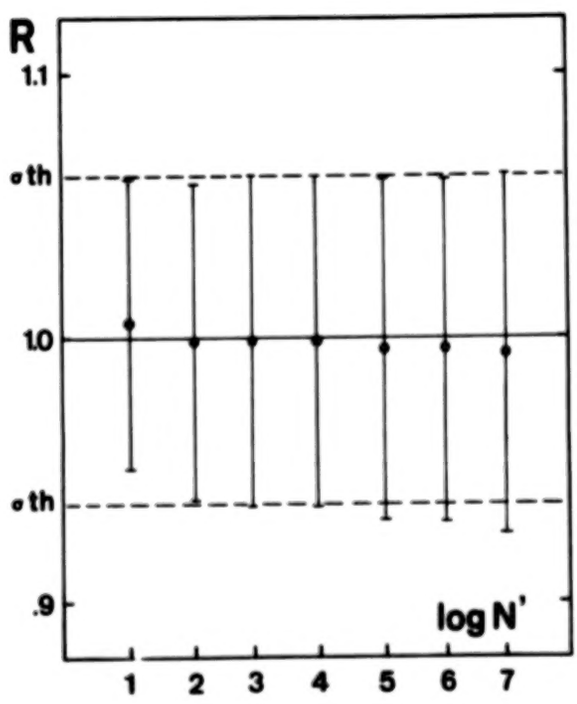


Fig. 8

Mean value of observed R as a function of log N'. The experimental conditions are as for Fig. 7. This plot shows that R is totally independent of N' over a 10^6 dynamic range. This also confirms that R does not depend on the scintillation. For the largest counting rates, the small decrease of R reflects its sensitivity to very small trends or possibly to autocorrelation in the scintillation. The observed standard deviation always agrees with the theoretical one.

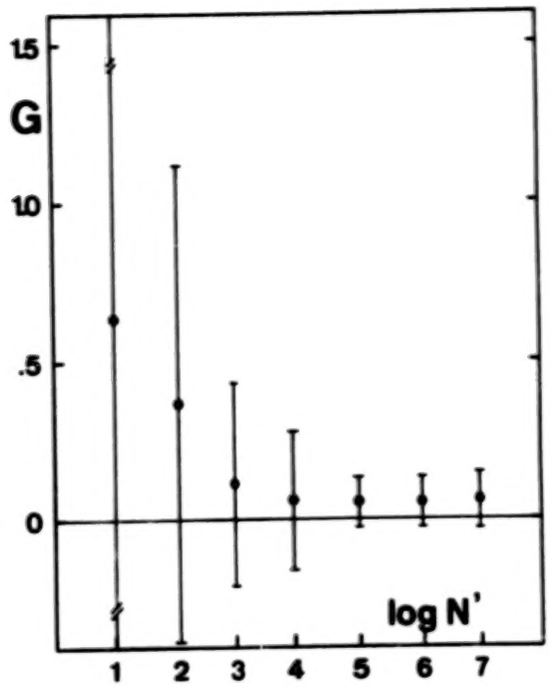


Fig. 9

Mean value of observed G as a function of log N'. The experimental conditions are as for Fig. 7. For N'<1000 cps, the plot shows the effect of small spikes. Only few extra counts cause the G values to grow. For N'>1000 cps, the small positive value of G indicates a small asymmetry in the scintillation.

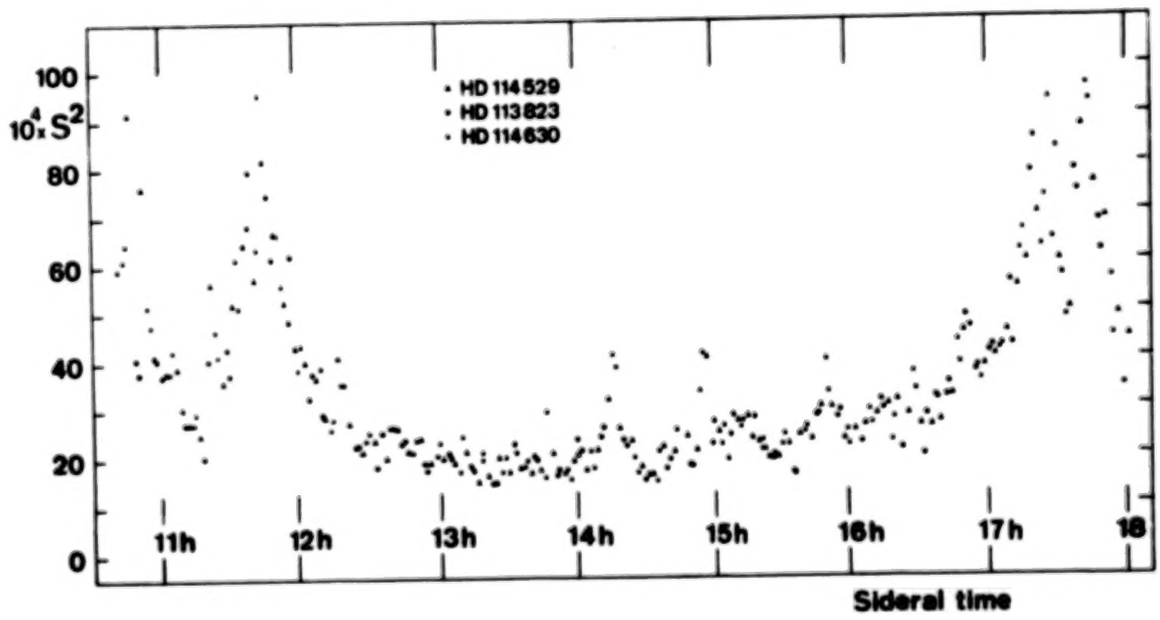


Fig. 10

This plot shows the variation of the scintillation estimator s^2 as a function of time during eight hours of observation. Each point is an average value computed over one minute for the seven filters. Three stars of mv = 4.56, 5.97 and 6.19 with spectral type B8, B9 and G0 were alternately observed. The points of each individual star are well mixed. The telescope aperture was then .4 m.

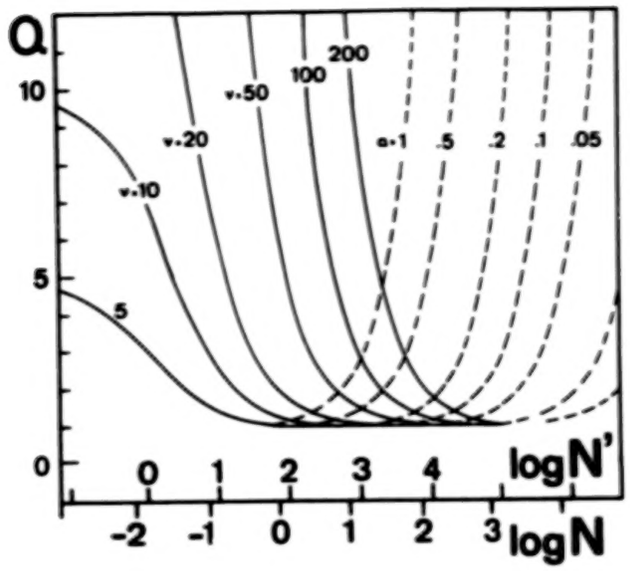


Fig. 11

Theoretical value of Q as a function of log N' for a single spike (or drop) of v counts occurring in one of the 256 samples (compare with Fig. 13). The broken lines represent the theoretical value of Q as a function of log N' for various linear trends between $(1-\alpha/2)N$ and $(1+\alpha/2)N$. Scintillation is neglected.

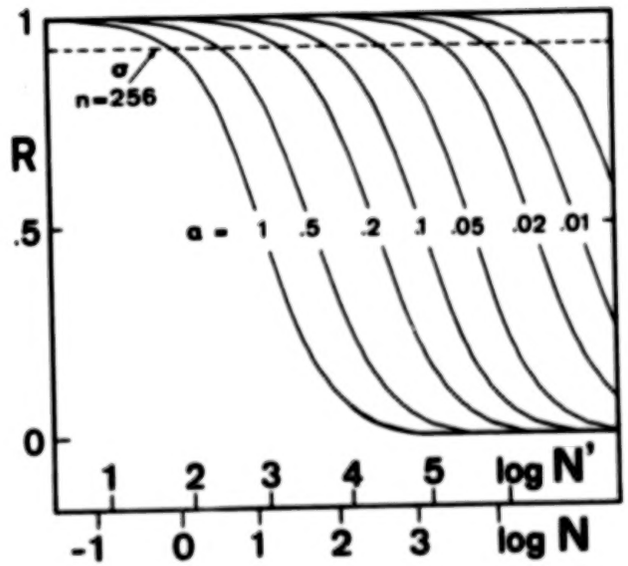


Fig. 12

Theoretical values of R as a function of log N' for various values of a linear trend between $(1-\alpha/s)N$ and $(1+\alpha/2)N$. The broken line is set at one $\sigma(R)$ for integration with n = 256 sub-intervals. Remember that the counting rate $N'=Nn/T$, N being the average count per elementary sub-interval.

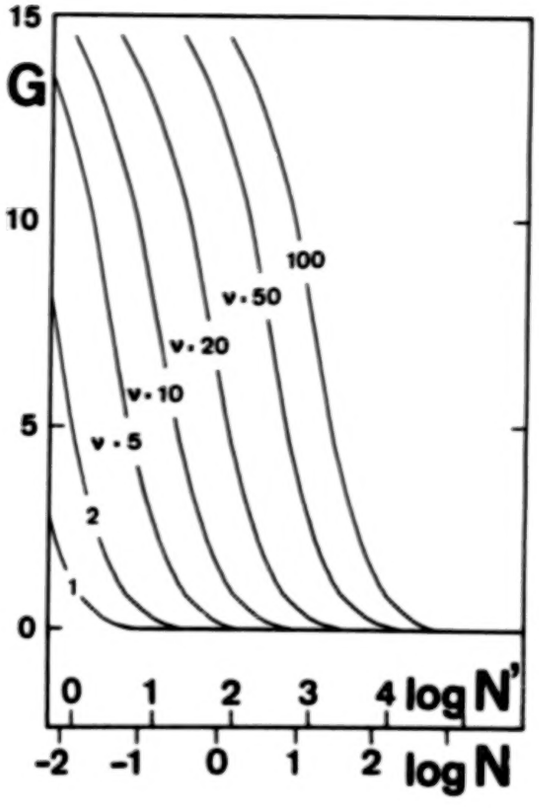


Fig. 13

Theoretical value of G as a function of log N' for a single spike (or drop) of v counts occurring in one of the 256 samples (compare with Fig. 12).

III. DETECTORS

D11

N85-17903

GAIN INSTABILITIES IN PHOTOMULTIPLIERS: HOW ACCURATE ARE PHOTON COUNTING MEASUREMENTS?

W. A. Rosen and F. R. Chromey
Vassar College

ABSTRACT

Experiments performed on five commercially available photomultiplier tubes indicate that gain instabilities can be an important source of error in photon counting measurements at the 1% level. We show that the error cannot be significantly reduced by standard differential measurement techniques. Analysis of time variations in the pulse height distribution is shown to be a sensitive diagnostic tool for the measurement of gain variations. Using this technique we find that gain variations occur at counting rates as low as 100 Hz. We argue that such errors will be present at some level in all tubes and discuss several calibration schemes capable of reducing the error to below the 0.1% level.

INTRODUCTION

Gain instabilities in photomultipliers have been observed almost since the time of their development (Marshall et al. 1947). Gain variations are generally divided into two categories – long term drift (with time scales of days to months and years), generally associated with aging of the tube and which is often irreversible; and short term drift (seconds to hours), traditionally known as "fatigue." The term "fatigue" is misleading, since the gain in many tubes increases following an increase in incident light. In some tubes the gain can change first in one direction and then in the other following a change in incident light level. The gain variation tends to be very rapid at first, levelling off to a few percent over periods of hours. When the light source is removed, the tube generally recovers, though not necessarily to its original state (Youngbluth 1970). The effect evidently depends in a complicated way on the recent history of illumination, making any simple calibration scheme impossible. The magnitude of the effect can vary widely even among tubes of the same manufacturer's type, so that measurements made on one tube may have little applicability to other tubes of the same model. There is a great deal of evidence which suggests that all tubes fatigue to some extent at all incident light levels (see e.g. Sommer 1973, Coates 1975).

The mechanisms which produce gain instabilities are not yet well understood. Most authors agree that the problem lies in the electron multiplier (see e.g. Youngbluth 1970), however even the exact physical location within the multiplier is not known. While one might suspect that the effect occurs low down in the multiplier chain where the current is highest, Youngbluth (1970) showed that, in five S-11 tubes, the earlier dynodes contributed as much as the later stages, and that, in fact, the last stage was the most stable of all. Coates (1975) also showed that, in several RCA GaP and EMI venetian blind tubes, a component of the effect was due to the earlier stages.

At least three different models have been proposed as the source of gain instabilities. The oldest is due to Marshall et al. (1947). This model involves the removal of cesium from the dynode surfaces by electron bombardment. Since the electron gain of a given stage is a sharply peaked function of the amount of cesium present, the gain can either

increase or decrease depending on the amount of cesium on the dynode before the change in incident light. When the light is removed, the cesium returns to more or less its original state. A second model involves the charging of the dynode insulators (Youngbluth 1970). This mechanism may be responsible for the initial rapid component of the gain variation, particularly in venetian blind type tubes (Boutot et al. 1983). The third model involves the formation of a thin oxide layer on the dynode surface. Cantarell and Almodovar (1963) have suggested that the gain variations are due to the presence of the Malter effect in this layer, and Yamashita (1980) has proposed that the variations are due to the trapping of electrons and holes within the layer. This model is at least consistent with the results of Koosman (1964) who found that the magnitude of the variations seemed to correlate well with the amount of oxygen present in the tube. In view of the complicated behavior of the effect it is likely that more than one mechanism may be involved at the same time.

Over the years there have been many unsuccessful attempts to "cure" the problem of gain variations. Typically these involve preflashing the tube with a bright light source, varying the high voltage, and distributing the high voltage asymmetrically among the dynodes. One manufacturer (Hamamatsu) coats the ends of the electrodes with chromium and maintains them at the cathode potential to minimize charging effects.

Coates (1975) has proposed a scheme in which the gain changes are calibrated by renormalizing the pulse height distribution of the tube's output pulses. Lopez and Rebolledo (1981) have tried chopping the incident light with duty cycles comparable to the time required for the recovery of the tube, but found that the recovery time in most tubes was too long to make this practical. Several calibration schemes have been described which employ a calibration standard which is turned on periodically (Ageno and Felici 1963, Black and Valentine 1964, Comunetti 1965, and Rosen 1982). In typical astronomical applications these methods tend to be sensitive to interference from noise components in the incident light near the frequency of the reference signal.

The effects of gain instabilities can be minimized (but not eliminated) by using the tube in the photon counting rather than the current mode, with the degree of improvement depending on the setting of the low level discriminator. Fatigue in pulse counting measurements has been observed by several authors. Coates (1975) found gain changes of several percent in two RCA GaP (types 8850 and 8852) and three EMI venetian blind (types 9558, 9658, and 9659) tubes. He also found that in some tubes the level of fatigue increased with decreasing temperature. Yamashita (1978, 1980) observed a hysteresis effect in the gain of two Hamamatsu R580 tubes when the the tube was cycled through a range of incident light levels. He found that the gain change could be well represented by a sum of three exponential functions with different time constants. Lopez and Rebolledo (1981) found that an RCA C31034 photomultiplier fatigued by as much as 5% over one hour when exposed to counting rates up to 1 Mhz.

In this paper we investigate the magnitude of the error which may be introduced by photomultiplier gain instabilities in typical photon counting photometry. We argue that many of the standard calibration methods and tests for gain variations are not sufficient to guarantee accuracy at levels better than 1%. Finally we show that analysis of time variations in the pulse height distribution can provide a powerful tool for diagnosing gain instabilities and discuss several apporaches which appear capable of stabilizing tube gain at a level of accuracy at least an order of magnitude better than those now available.

EXPERIMENTAL METHOD

The experimental apparatus has been described in detail elsewhere (Rosen and Chromey 1984). The photomultiplier tube to be tested is placed inside a light tight temperature

controlled box and separated by a shutter and neutral density filters from an LED light source (see Figure 1). The output current pulses from the tube are converted to voltage, amplified, shaped (OEI AH0008 preamp and EG&G 572 pulse shaping amplifier) and sent to a multi-channel analyzer (Tracor- Northern TN-1705 or TN-7200) with on board low level discriminator. The analyzer could be used either as a multi-channel scaler (MCS) to measure counting rate vs. time; or as a pulse height analyzer (MCA) to give a histogram of the pulse height distribution. For data storage, the multi-channel analyzer was interfaced to a microcomputer (Tektronix 4051). The LED intensity, box temperature, and tube high voltage power supply were monitored continuously. Tests were performed over a range of operating conditions to establish the effects of any observed variations upon the photomultiplier output counting rate.

All of the tests performed were repeated at least twice, and the more critical tests (particularly at low counting rates) were repeated many times. During most of the testing the tube was left on continuously at full voltage. In addition, several of the tests were repeated with the tube voltage turned down to 500 volts between tests.

Five comercially available photomultipliers were tested, an ITT FW-130, two EMI 9502S, an RCA 8575, and a Hamamatsu R647. The first tube is an S-20 box and grid type. The second type is an S-11 venetian blind, and the last two are bialkali linear focused. All the tubes tested were found to fatigue at about the same level of magnitude, with the Hamamatsu tube performing marginally better than the rest. There was no obvious difference in the performance of the two EMI tubes. The most extensive testing was performed on the FW-130, due to its widespread use at many observatories, and the results of these tests are presented in detail below.

RESULTS

Three basic types of tests were performed on each tube. First, a series of baseline tests were run, consisting of simply opening the shutter to expose the tube to a constant light source after the tube had been left in darkness for at least 18 hours. Data were acquired for a total of two hours in one minute sampling intervals. Dark counting rates were measured for several minutes before and after each test. The response of the ITT FW-130 at room temperature to such a test is shown in Figure 2. The change in counting rate vs. time following the opening of the shutter is plotted at counting rates of 1, 10, and 100 Khz. The tube is seen to fatigue in the positive sense. The magnitude of the change varies from about 6% at 100 Khz after two hours to about 1% at 1 Khz. The change is most rapid at first, and the time constant for the effect is inversely rate dependent. Upon repetition of the tests the curves did not reproduce exactly, but the sense and magnitude of the change remained nearly constant, supporting the notion that the tube can recover to more than one internal state.

The second type of test was performed in essentially the same way as the first except that the pulse height distribution rather than the counting rate was monitored periodically during the two hours. These tests demonstrate the process by which a D.C. gain variation produces a change in the counting rate, as can be seen in Figure 3. Figure 3a shows the pulse height distribution at 10 Khz immediately after shutter opening (solid line), and 120 minutes after shutter opening (broken line). Figure 3b shows the difference in counts vs. pulse height between the initial pulse height distribution and the distribution measured at 30 minute intervals. As the D.C. gain of the tube gradually increases, there is a corresponding shift in the distribution from lower to higher pulse heights. Moreover there is a shift in the height of pulses which originally were below the low level discriminator (indicated by a vertical bar) to the region above the discriminator, so that these pulses are now counted. Detailed analysis of the pulse height distributions reveals that, to within

the statistical error, the number of pulses shifted above the low level discriminator is just the number required to produce the changes in counting rate observed in Figure 2.

A number of tests were repeated at 0° C. These showed the interesting property that the fatigue initially occurred more rapidly and in the negative sense. Moreover the magnitude of the effect was about twice that at room temperature. In addition, a number of tests were run in which the tube voltage was turned down to 500 v. overnight, then brought to operating voltage for two hours before the shutter opening. These tests also showed the property that the fatigue initially occurred more rapidly and in the negative sense. This suggests that the gain variations are due to at least two mechanisms with different time constants, and that the recovery from at least one of these mechanisms is partly suppressed by the large dark current at room temperature.

The reduced dark count at lower temperature does not completely account for the increased gain variation. The low temperature pulse height distribution of the dark counts was found to be shifted in the direction of lower pulse heights. When a counting rate comparable to the room temperature dark count was introduced using a faint light source, the pulse height distribution was partially restored to its shape at room temperature, but a residual component remained, apparently due purely to temperature effects. This confirms the results of Coates (1975) and suggests that tubes which tend to fatigue in the negative sense and have low room temperature dark current may benefit from cooling.

A close examination of Figure 4 reveals that the variations in pulse height distribution are much larger than the corresponding changes in counting rate, since only a few counts are actually shifted across the low level discriminator. This suggests that analysis of the pulse height distribution can provide a sensitive diagnostic for the presence of gain variations. For example, if one calculates the ratio of the number of pulses above bin 50 (roughly 50% of the counts) to the total number of counts, this ratio changes by a factor 70% greater than the corresponding change in the counting rate.

The question arises of whether gain variations occur even at very low counting rates. The difficulty with answering this question lies in the fact that at low counting rates any gain changes can be expected to be small, while the counting time required to achieve high accuracy is large. For example, to achieve a statistical accuracy of 0.1% at 100 Hz, nearly six hours of sampling are necessary to obtain two data points. Fortunately the analysis of variations in the pulse height distribution discussed above provides a way out of this difficulty. Measurements of the pulse height distribution performed at 100 Hz over four hours (with the tube cooled to minimize the dark counts) show the same kind of pathological changes as the 10 Khz tests shown in Figure 3. Using the methods described below, we estimate that the gain changes by about 0.3% over this time period. The magnitude of the observed variation is smaller, and the time required to produce it is longer when compared to higher counting rates. However, since a much longer time is required to achieve statistical significance we conclude that simply reducing the incident light to low levels will not rescue us from the problem of gain variations. We further conclude from these measurements that such variations will probably occur at any useable counting rate.

The final series of tests were designed to determine the response of the photomultiplier to a changing incident light level. Figure 4a shows a sequence in which the tube is alternately exposed for two minutes to a fainter, then brighter light source, with two minute dark intervals in between. The effect of this test is to simulate a measurement of a faint "unknown" source followed by a brighter "reference" source. The measurement of the faint source is then "corrected" by dividing its counting rate by the rate of the following "reference". The change in this ratio over a two hour period for two different ratios (1:10 and 1:1) is plotted in Figure 4b. Evidently the complicated processes of fatigue and

recovery during a changing incident light input are sufficient to produce an error in the corrected measurements at about the 1% level. Moreover, no great improvement is gained by using a reference source whose intensity is near that of the unknown. During an actual observing run, such variations might easily be attributed to such effects as fluctuations in sky brightness or transparancy.

This correction procedure represents a simplified version of the correction methods employed in practical photometry. However, during actual measurements in an observatory, the relative spacing, duration, and relative intensities involved in the measurement sequence are much less regular, and often not recorded. Therefore we would expect that the errors introduced into an actual measurement by fatigue effects are at least comparable to those seen in Figure 4.

CURRENT RESEARCH

At present we are investigating two possible avenues of calibrating the effective gain. One of these involves chopping the incident light at high speed alternately between the light source to be measured (unknown) and a standard light source whose intensity is monitored by an accurate photodiode. The component of the standard source which is incident on the tube is highly attenuated by neutral density filters, so that the photodiode sees a bright and easily measureable source while the photomultiplier sees a standard which is about several times the intensity of the unknown. The counting rates of the unknown and standard are measured independently and the former divided by the latter to correct for gain variations. The intensity and duty cycle of the standard signal are chosen to minimize statistical errors. Preliminary baseline measurements performed with this method indicate that is is easily capable of accuracies better than 0.1%. The fundamental limitation is the gain linearity of the tube over the range of standard to unknown intensities. This method has the advantage that the tube is always looking at a bright source, even when no unknown is present. therefore the tube is, in a sense, prefatigued, thereby eliminating the rapid component of the gain variation. The method has the added advantage that it automatically corrects for any variation in the power supply voltage or amplifier gain. For example, changing the power supply voltage by one volt produced a noticeable change in the counting rates of both the unknown and standard, but no noticeable change in the ratio of the two.

We are also investigating two variations of this method. In one, the signal from the standard is not alternated with the unknown, but rather is simply periodically added to it, and then digitally separated form the tube's output signal. This has the advantage that it eliminates the mechanical chopper, but it may suffer from the problem of interference from noise components in the unknown signal. In the second variation, the standard signal is controlled by a feedback circuit so that its resulting counting rate matches that of the unknown. The signal from the photodiode is then a fixed multiple of the unknown. The photomultiplier acts simply as an optical multiplier for the diode. While this method greatly increases the counting time required for a given statistical accuracy it has the advantage that the tube is looking at essentially the same signal, and therefore does not rely on the linearity of the tube over a large dynamic range. At this preliminary stage of testing it is difficult to say which of these variations is best suited for the complicated behavior involved in gain variations.

An alternative method which does not suffer from the problems mentioned above is suggested by the analysis of the pulse height distribution. Since the changes in pulse height distribution are much greater than the observed changes in the counting rate, these changes can provide a powerful diagnostic for estimating and correcting for gain variations.

If the shape of the pulse height distribution is known, it is easy to predict ΔN, the change in the counting rate that will result from a given change in the tube gain, ΔG. Coates (1975) shows that for small values of $\Delta G/G$:

$$(1) \quad \Delta N/N = -z_1 S(z_1)\Delta G/(G\alpha(z_1)) = -K(z_1)\Delta G/G.$$

Here z is the normalized pulse height, h/h_0, and z_1 is the setting of the low level discriminator. $S(z)$ is the probability density distribution for the pulse heights and $\alpha(z_1)$ is the counting efficiency:

$$\alpha(z_1) = \int_{z_1}^{\infty} S(z)\,dz.$$

Thus the total number of photomultiplier pulses is $N_0 = N/\alpha(z_1)$. The function $K(z)$ starts at $K(0) = 0$ and increases monotonically with z. Coates (1975) used equation (1) to estimate corrections for gain changes in a pulse counting photometer. He monitored the anode current of the photomultiplier, which is much more sensitive to ΔG than is N, and was able to obtain corrected $N(z)$ to 0.07% over 25 minutes.

We suggest here a method for monitoring gain changes which is based entirely on measurements made in the pulse counting mode. This method is similar to the one suggested by Coates, but has the advantage that no measurements in the current mode need be made, so that it should be both simpler to implement and more accurate at low counting rates. Consider an apparatus constructed so that counts can be accumulated through two different low level discriminators. If the first discriminator is set at z_1 in the usual way (i.e., as low as possible: see Young 1969), the counting rate it registers will be fairly insensitive to gain changes since $K(z_1)$ will be close to zero. If the second discriminator is set at $z_2 >> z_1$, then the counting rate it measures will be much more sensitive to gain changes. The ratio of the two rates, $R = N(z_2)/N(z_1)$ changes with gain according to:

$$\Delta R/R = (K(z_2) - K(z_1))\Delta G/G.$$

so that the gain related changes in $N(z_1)$ will be

$$(2) \qquad \Delta N(z_1)/N(z_1) = (1 - K(z_2)/K(z_1))^{-1}\Delta R/R = k_{12}\Delta R/R.$$

The factor k_{12} will be known if the normalized pulse height distribution, $S(z)$, is known. For example, if we choose z_2 so that R is about 2, the observed distribution shown in Figure 5 gives a value of 0.07 for k_{12}.

Empirically, we find that $S(z)$ for a fatiguing tube is invariant over short times periods (hours). Figure 5; for example, shows $S(z)$ for the two pulse height distributions shown in Figure 3a. Any differences between the two curves in Figure 5 are within the observational scatter (about 1% near the peak of the distribution). $S(z)$, however, does change over longer time scales, and the value of k_{12} must be redetermined periodically. If a multi-channel analyzer is not available, k_{12} may be determined simply by setting narrow discriminator windows at z_1 and z_2. For a 10 khz test like that in Figure 3 the total uncertainty in the right hand side of equation (2) is about 10%, if z_2 is selected to be near 1. This uncertainty is due mainly to uncertainties about the invariance of $S(z)$. Since $\Delta N(z_1)/N(z_1)$ is typically 1.5% over a four hour test, one could expect corrected counting rates that are accurate to 0.15% over four hours using this scheme.

Implementation of this method in a pratical gain-insensitive photometer would require a simple microcomputer to sense R and correct $N(z_1)$ in real time, or at least a data acquisition system that would separately record $N(z_1)$ and $N(z_2)$ so that the correction

could be made at a later stage. This method does require both a knowledge of $S(z)$ and some assurance of its stability. If the shape of the pulse height distribution for dark counts is very different from that of the signal counts, then significant variations in the ratio of dark to signal counts will produce variations in $S(z)$. We find that this is not the case for the tubes tested here, but note that it may be a problem in some circumstances (Young 1969). This problem could be corrected if the pulse height distribution for the dark counts is known. In general, if the shape of the pulse height distribution can be monitored for changes, then the effects of gain instabilities can be corrected to a few tenths of a percent. In principle, the method is limited only by the usual counting statistics and by an imperfect knowledge of the pulse height distribution .

CONCLUSIONS

These results suggest that gain instabilities in photomultipliers may be an important source of error in pulse counting photometry at the 1% level. Until the exact mechanisms responsible for the effect are known, none of the correction methods such as preflashing, chopping the incident light, or varying the high voltage are likely to be effective or reliable. However the two calibration schemes described above are capable of reducing the effects of multiplier gain instabilities to the 0.1% level. Both methods have the virtue of also correcting for variations in power supply voltage and in amplifier gain. In addition, the method based on the pulse height distribution can correct for long term drift in the tube if the function $S(z)$ is remeasured periodically.

REFERENCES

Ageno, M., and Felici, C. (1963). Rev. Sci. Instrum. **34**, 997.
Black, J.L., and Valentine, E. (1964). Nuc. Instrum. Methods **31** , 325.
Boutot, J.P., Nussli, J., and Vallat, D. (1983). in Advances in Electronics and Electron Physics, vol. 60, edited by P.W. Hawkes (Academic Press, New York), p. 223.
Cantarell, I., and Almodóvar, I. (1963). Nuc.Instrum. Methods **24** , 353.
Coates, P.B. (1975). J. Phys. E **8**, 189.
Comunetti, A.M. (1965). Nuc. Instrum. Methods **37**, 125.
Hamamatsu Photomultiplier Tube Cat. No. SC-1-4-1, p. 8.
Koosman, J.G. (1964). IEEE Trans. Nucl. Sci. NS – 11, No.3 , 56.
López, R.J., and Rebolledo, M.A. (1981). Rev. Sci. Instrum. **52** , 1852.
Marshall, F-H, Coltman, J.W., and Hunter, L.P. (1947). Rev. Sci. Instrum. **18**, 504.
Rosen, W.A. (1982). B.A.A.S. **14**,646.
Rosen, W.A., and Chromey, F.R. (1984). submitted to A.J.
Sommer, A.H. (1973). App. Optics **12**, 90.
Yamashita, M. (1978). Rev. Sci. Instrum. **49**, 1336.
Yamashita, M. (1980). Rev. Sci. Instrum. **51**, 768.
Young, A. T. (1969). App. Optics **8**, 2431.
Youngbluth, O. (1970). App. Optics **9**, 321.

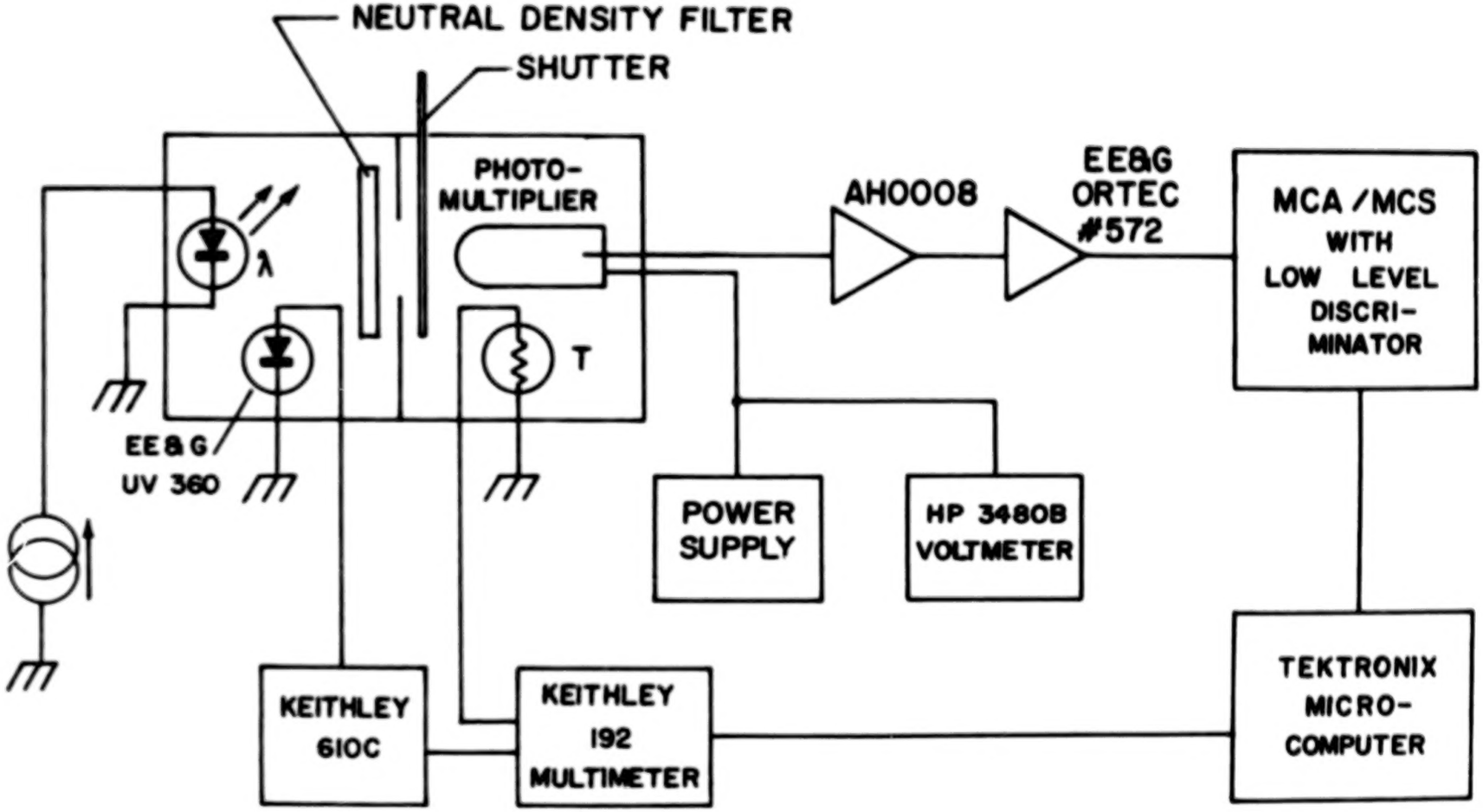


Figure 1.- Schematic of the experimental apparatus. The box temperature is monitored by thermistor, T. The LED ouput is monitored by the EG&G UV-360.

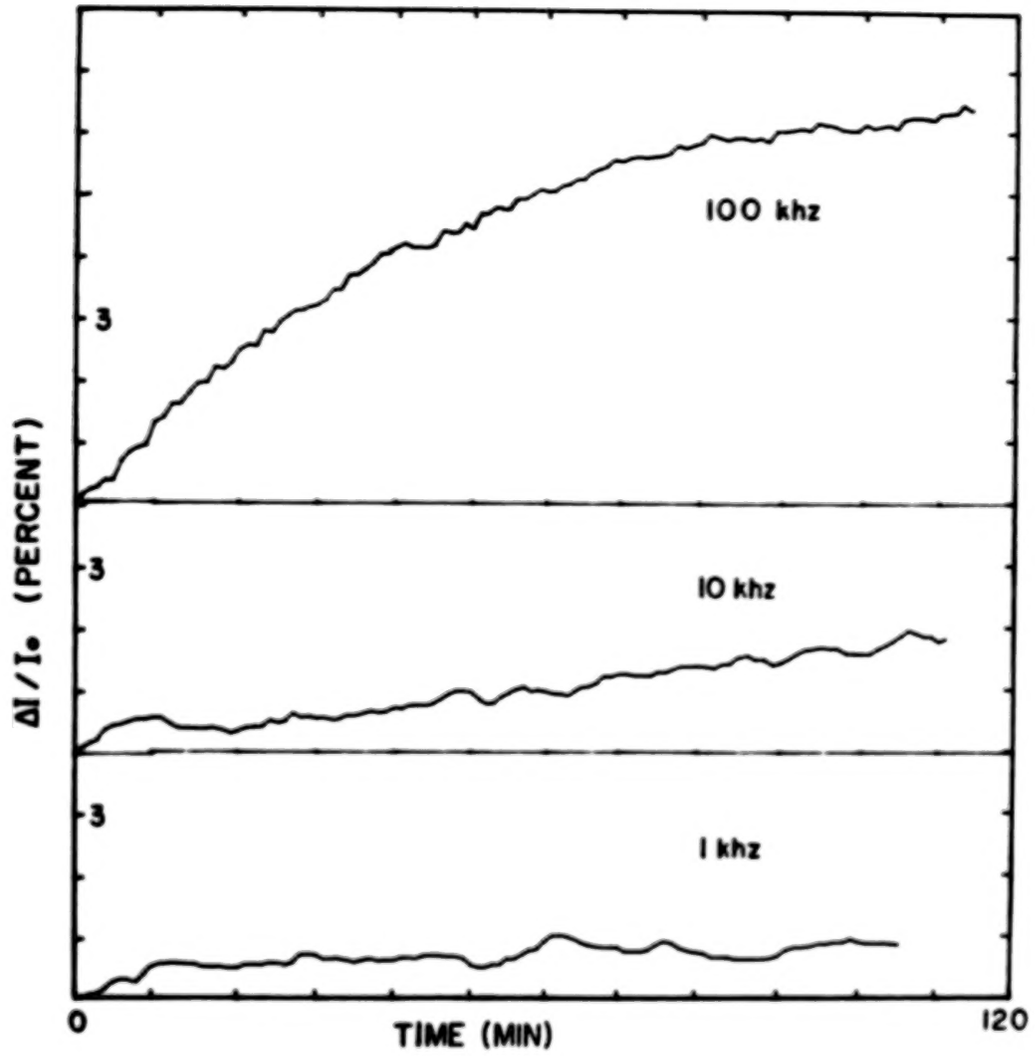


Figure 2.- Results of the baseline test for the ITT FW-130. The relative change in counting rate as a function of time after the shutter opening is given for counting rates of 1, 10 and 100 Khz. Counts were acquired in one minute sampling intervals. The lower two curves are smoothed by 4 and 10 point running averages respectively, and the r.m.s. noise is (top to bottom) 0.04, 0.06, and 0.13%.

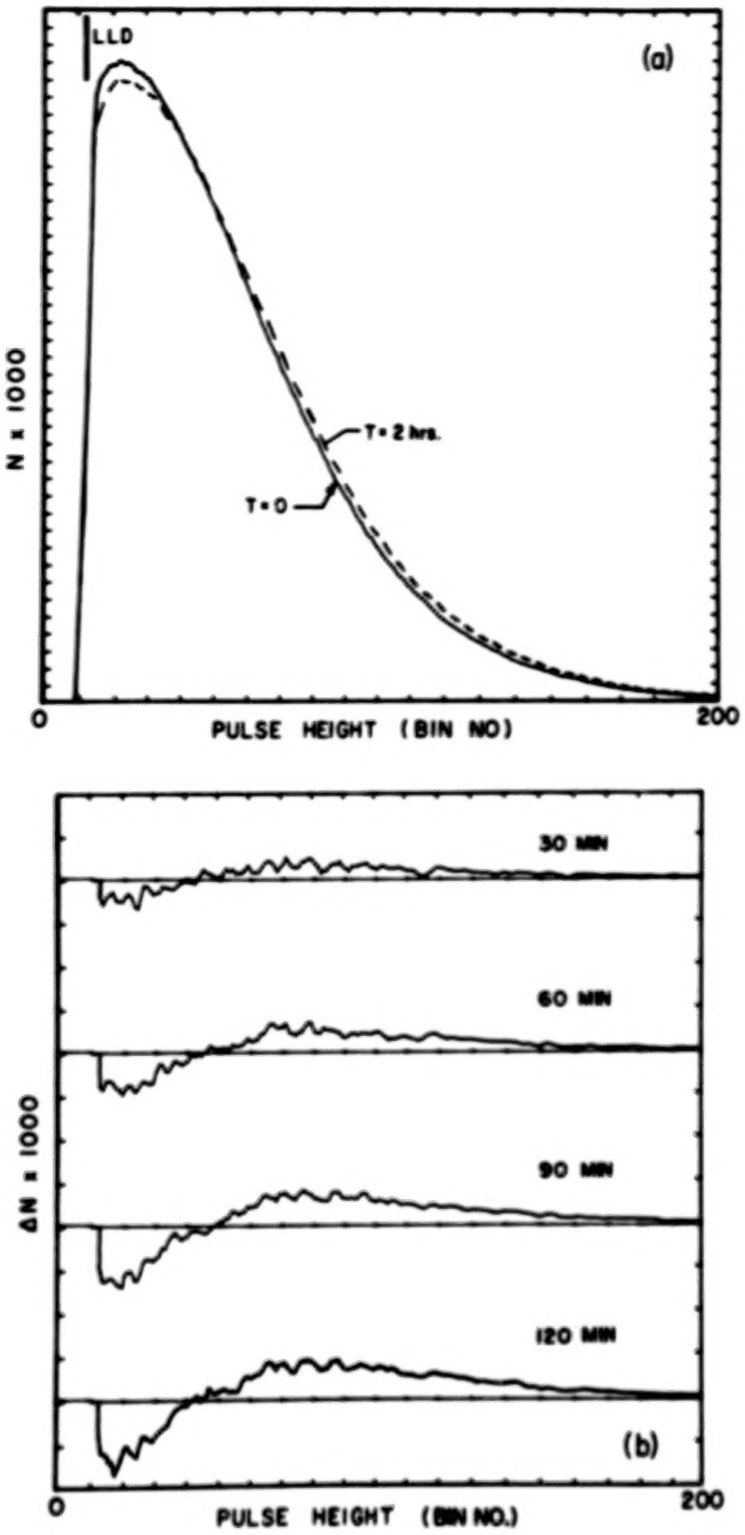


Figure 3.- Part (a) shows the pulse height distribution in the 10 Khz baseline test, as measured at the start of the test, and again at two hours after shutter opening. Part (b) shows the change from the initial pulse height distribution at four different times after shutter opening.

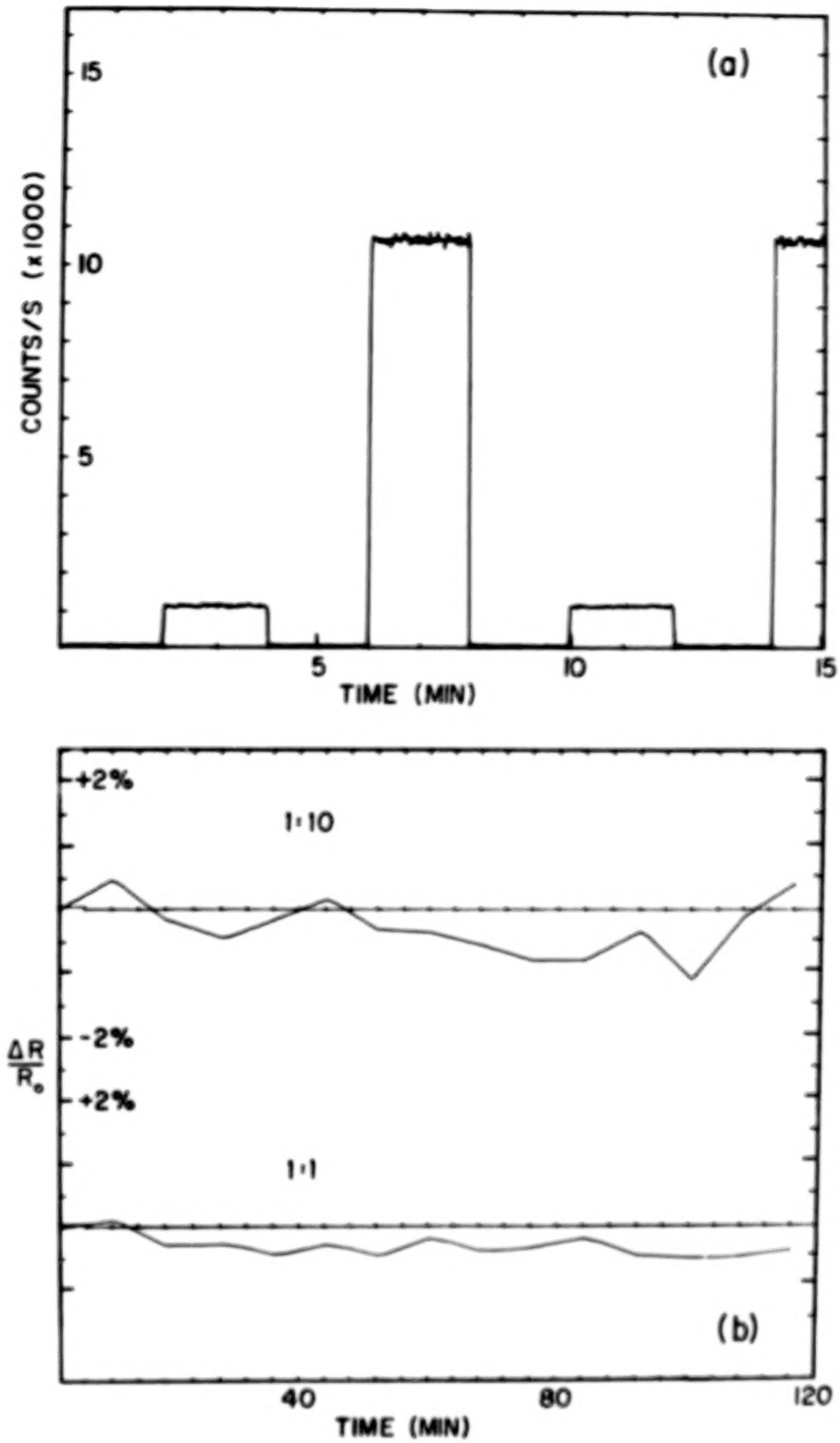


Figure 4.- Part (a) shows the counting rate as a function of time in the test that alternated between "unknown" and "reference" light levels. The change in the ratio of unknown to reference counts is shown in part (b). Note the different time axes for (a) and (b).

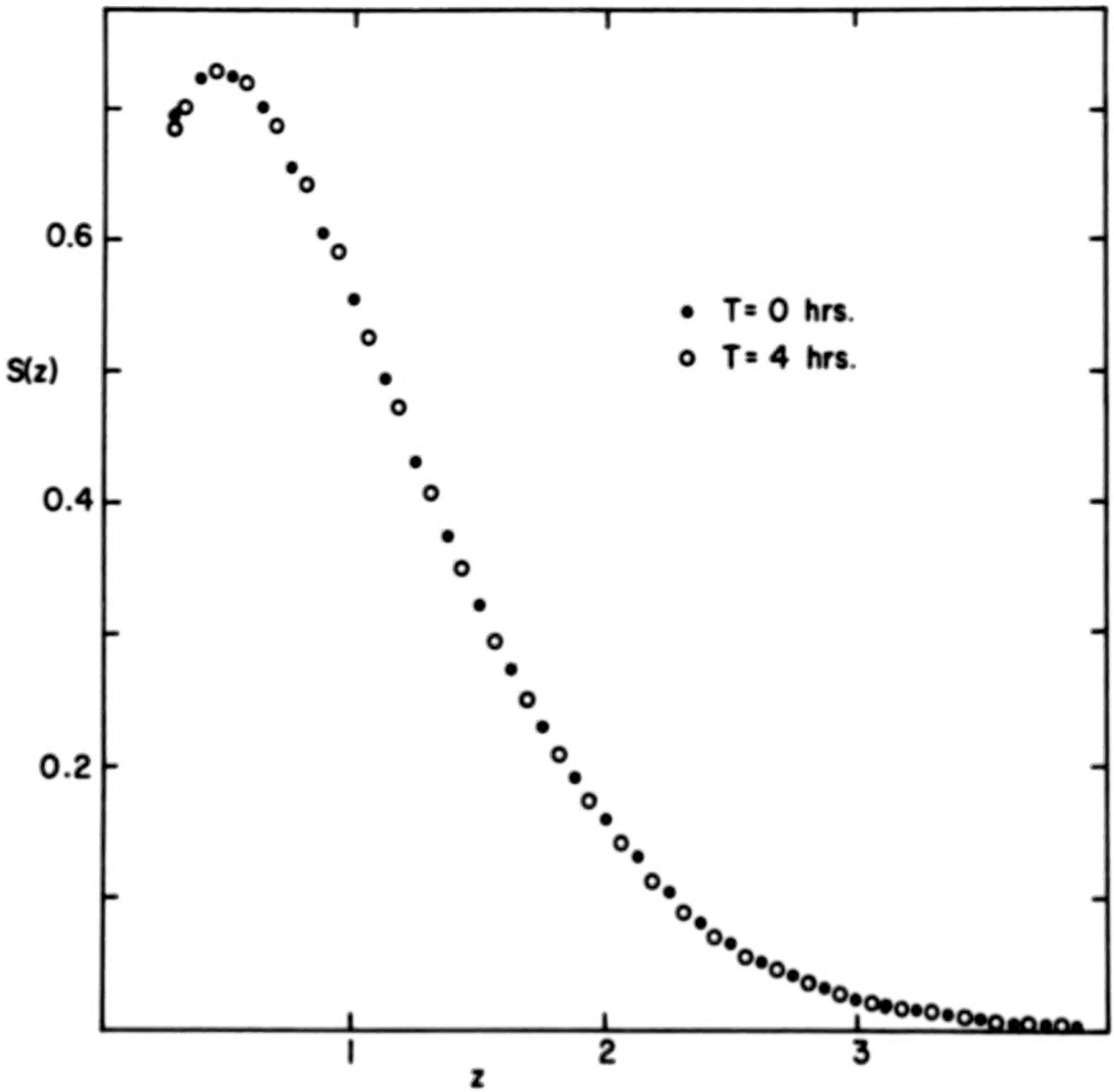


Figure 5.- The probability density distribution of the normalized pulse height, z. Data are shown for the two raw distributions pictured in figure 3a.

D12

N85-17904

A CCD SYSTEM FOR PHOTOMETRY OF DIRECT AND SPECTROSCOPIC IMAGES

J. E. Frecker, T. Gehrels, R. S. McMillan, W. J. Merline,

M. L. Perry, J. V. Scotti, and P. H. Smith

Lunar and Planetary Laboratory, University of Arizona, Tucson AZ 85721

I. Summary/Introduction

We are using a charge-coupled device (CCD) for direct imagery at the focus of a ground-based telescope and for high dispersion spectroscopy of stars. The CCD is an RCA SID 53612 thinned, buried channel array of 512x320 30-micron square pixels that are back-illuminated and refrigerated to about -60 C in a rugged, RF-shielded vacuum housing. The double-correlated sampling (integrator) readout system, the Z80 microprocessor readout control system, and the data collection system are described. The readout noise is 120 electron-hole pairs (ehp) per pixel per readout, the thermal dark current is 50 ehp/pixel/sec, and the scale of the 12-bit analog-to-digital converter is 25 ehp/AD unit. Unique features of our system compared to other astronomical implementations of the same model of CCD are the fact that the CCD chip was not preselected for good performance in scientific application, the use of dry ice as a refrigerant, which greatly simplifies the problem of temperature regulation, an unusually simple but versatile clocking and readout control system, and the extensive use of the system in two unusual modes of optical imaging. For the purposes of this Workshop, our system and our experience with it are presented in the form that will be most useful to those planning to do short-term, accurate differential photometry of two or more "bright" stars or of several spectral bandpasses simultaneously. Performance of most other astronomical CCD systems is usually only quoted for low light levels (limiting magnitude, etc.). Although our system was optimized for simplicity, reliability, and low cost, instead of for accurate differential photometry, the accuracy of differential photometry of bright laboratory light sources can be as good as 1 part in 700. Similar photometry of "bright" astronomical sources is limited by the behavior of the atmosphere. We conclude that a CCD system optimized for accurate short-term differential photometry at "high" light levels could provide a ratio of signal to noise > 1000, neglecting external sources of noise, and we recommend some features of such a system.

II. Description of Hardware

II.A. CCD chip

We have an RCA model SID 53612 thinned, buried channel CCD with 512 x 320 pixels. The currently available equivalent of this CCD is the RCA model SID 501EX (some of which we have also recently purchased). The pixels are 30 micron squares with no dead space between the pixels. The quantum efficiency of this type of CCD has been measured by Fowler et al. (1981). It is one of the largest scientific CCD's readily available by routine purchase order, and has been put into astronomical use by several groups, including the Kitt Peak

National Observatory, Photometrics, Ltd. of Tucson, and Lick Observatory. However, most institutions carefully select the actual chip they use by purchasing a large number of the same model, and testing them all. Thus the penurious astronomer desiring his own CCD system cannot rely on the published performance figures for CCD's put into operation by large observatories to describe what he can expect from a single CCD ordered and selected at random. Our CCD is such an unselected SID 53612, so the performance we quote comes closer to the average.

II.B. CCD Control and Readout System

To control the timing of charge transfer and readout we use an 8-bit Z80 microprocessor-based system with a console printer, oscilloscope display, and sufficient digital memory to save all of the most recent CCD "picture", for display purposes. See Figure 1. We control the subpixel timing by downloading bit patterns from the Z80's main memory into a special 4 kilobyte RAM, from which the bit patterns are transmitted as waveforms that control the chip. Since these clock-timing waveforms are created by software, the readout process can be altered easily to test concepts such as "binning", reading parts of the CCD selectively, and synchronizing the rate of transfer of charge to the rate at which an optical image drifts across the CCD (see below). To minimize the effects of RF interference and other external mechanisms of electronic noise, a preamplifier is mounted next to the CCD, inside the same RF shield, before the signal is transmitted to the external electronics. A conventional double correlated sampling (charge integrating) circuit, followed by a sample & hold and an analog-to-digital converter (A/D) complete the analog processing of the signal, giving 12 bits of digital information describing the amount of exposure each pixel received from the optical image.

II.C. Refrigeration Housing and RF Shield

The RF-shielded, light-tight refrigeration housing is shown in Figure 2. The CCD is kept in a moderate vacuum (less than 1 mm of pressure) to keep it dry when cold. A full charge of dry ice refrigerant cools the CCD to about -60 C and lasts about 8 hours. Dry ice is not the currently recommended cryogen for CCD's because the performance continues to improve as the chip is cooled, until about -100 to -140 C (Geary & Kent 1980). However, the systems that provide this cold an operating temperature use liquid nitrogen as a refrigerant, which is much colder than the desired operating temperature. Thus a sophisticated, active system to heat the CCD to a controlled temperature is needed with liquid nitrogen. Dry ice has the advantage that no active temperature regulation is required.

II.D. Data Recording System and Instrument Interface

Figure 3 is a schematic of the real-time data recording and display system, as interfaced to the radial velocity spectrometer described below. In real time, we can display a gray-scale video picture of the signal from the whole CCD array, and display or print out selected sections of data from the current observation or an earlier one read back from the tape. The number of CCD pictures that can be stored on one 1200-foot reel of tape at 800 bytes per inch depends on the mode of data compression; 112 compressed frames per reel is a typical number.

III. Performance of Hardware

III.A. Electronic Noise

The rms readout noise is +_ 120 electron-hole pairs (ehp) per pixel per readout and the thermal background from the CCD at approximately -60 C is 50 ehp per pixel per second. If cooled to -100 C, one can expect the thermal background to disappear for most practical purposes, and the readout noise to go down to about +_ 50 ehp/pixel/readout. A CCD system optimized for accurate differential photometry should be colder than we are operating ours, to reduce the noise to those levels. The cosmetic defects on our CCD can be seen in Figure 4; they are the bright pixels on the right side of the picture. They never change position or intensity, so we have been able to avoid the contaminated areas in all our data reduction software.

III.B. Direct Imaging

We use this system to survey the sky for new asteroids and distant comets at the Newtonian focus of the 0.91-meter telescope of the University of Arizona Observatories on Kitt Peak. This telescope has a focal ratio of f/5, giving an image scale of 44.8 arcseconds per mm or 1.34 arcseconds per CCD pixel. The paraboloid primary mirror allows a useful field diameter of at least 50 mm, which is more than adequate for this CCD, which has an image diagonal of 18 mm. The image scale is a good compromise between the need to detect the angular motion of solar system objects and the need to scan a sufficiently large solid angle per observing night to find new objects. McMillan and Stoll (1982) showed that a scale between 1 and 2 arcseconds per pixel was optimum for this type of work.

III.B.1. "Stare" Observations

A "stare" observation is a conventional one in which the telescope is driven at the sidereal rate while a mechanical shutter in front of the CCD is opened, the CCD is exposed to light without clocking (transferring charge), the shutter is closed, and the charges accumulated in the CCD pixels are then transferred ("read out") to the computer. With dry ice as a refrigerant, the longest exposure time we can use is about 10 minutes because in longer times the CCD pixels are filled by the thermal background current. In 5 minutes of exposure on the 0.91-meter telescope stars of the 21st visual magnitude are sufficiently above the noise to be detected.

III.B.2. Drift-scanning Observations

The frame-transfer architecture of this CCD is characterized by the fact that the charge packets are transferred through the array of detecting capacitors, instead of being read into registers concealed on the chip. This allows the chip to be read and exposed simultaneously and continuously for as long as the data collection system can store the accumulating scan. We take advantage of this property to scan a long, unbroken strip of sky while the telescope drive is off. The electronic "bucket brigade" transfers of pixel charge are clocked at the same rate at which the optical image of the sky drifts across the

stationary detector. The time it takes for a star image to traverse the 512 rows of the CCD is the effective exposure time; at the celestial equator this is about 46 seconds at our image scale. In this manner we can survey a large solid angle with a detector that is physically quite small, compared to a photographic plate. Figure 5 is a sample of a picture frame from a scan. It shows everything that appears on the Palomar Observatory Sky Survey print of the same region.

III.B.3. Performance in Direct Imaging Mode

On a clear, moonless night the combined effects of readout noise, dark current, and sky background result in a standard deviation of +_ 6 to 10 analog-to-digital units (ADU) (= 150 to 250 ehp) per pixel in a 46-second exposure. This applies to both "stares" and scans. In both the "stare" and scan modes, we have a limiting magnitude of 19.5 (visual) for a 46-second exposure. This limit is defined such that the peak intensity of a star image is 6 standard deviations above the sky noise. The signal rate for all the light in a star image (not just the peak pixel) is 11.3 ADU per 19th magnitude star. For "bright" (15th mag), unsaturated star images, differential photometry can repeat to an accuracy of +_ 3 %. Because we show below that the CCD is capable of much better accuracy, we believe much of this 3% error is due to the instability of the sizes, shapes, and positions of the stars imaged on the CCD.

III.C. Spectroscopic Imaging

III.C.1. Scientific Purpose

The goal of this project is to develop a system capable of measuring radial acceleration of solar-type stars with an accuracy of +_ 5 meters/second/year. The original form of the instrument was conceived and built by the late Krzysztof M. Serkowski. It has been redesigned and rebuilt and is now being used in trial observations of integrated-disk sunlight and bright stars.

III.C.2. Optical System

he LPL stellar accelerometer is an interferometrically calibrated spectrometer (see Figure 6). Light is fed from the telescope focal plane to the spectrometer entrance aperture via a single optical fiber. Wavelength calibration is imposed directly on the starlight by a Fabry-Perot interferometer used in transmission. This interferometer is in turn calibrated by a number of independent wavelength references. The spectrum is then dispersed using an echelle grating crossed with a conventional plane grating. The two-dimensional spectrum is focused on the above-mentioned CCD. The appearance of the spectra from several echelle orders, modulated by the transmission profile of the Fabry-Perot interferometer, is shown is Figures 4 and 7. Each "spot" corresponds to one order of interferometric transmission through the etalon. The relative brightnesses of the spots form a set of samples of the intensity of the stellar spectrum in 0.05 Angstrom bandpasses, separated by 0.63 Angstroms. The optical fiber acts as an image scrambler. This means that changes in the image of the star in the telescope focal plane

result only in a change in brightness at the spectrometer entrance aperture and do not alter either the path of the light in the spectrograph or the distribution of light in the beam. The Fabry-Perot is a monolithic unit of fixed spacing, vacuum gapped, and employing tilt-tuning as the scanning mechanism. The etalon is stabilized at the temperature at which the coefficient of thermal expansion of the plate and spacer material crosses through zero.

III.C.3. Method of Operation

Since the tilt-tuning of the Fabry-Perot interferometer can be reproduced to an accuracy of 10 ppb between stellar observations separated by months or years, it is possible to reproduce the same wavelengths of maximum transmission through the etalon to within +/- 0.04 milliangstroms, or +/- 3 meters/second. If the radial velocity of the star has changed by (say) 10 meters/second between the times of the two observations in question, then the etalon will be transmitting light from slightly different parts of the intensity profile of the stellar spectrum. The ratios of the intensities of the images associated with the various etalon transmission maxima will change by small amounts, depending on the slope of the stellar spectrum at each spectral point observed. It is these small changes in relative intensities that we observe to detect small changes in Doppler shift. In practice, the etalon tilt is not reset to the same value for every observation of a star, but is adjusted (up to +/- 33 km/s) to compensate for the changes of the velocity of the spectrometer along the line of sight to the star and any small, long-term variations in the physical spacing of the etalon plates.

III.C.4. Ratios of Photometric Signal to Noise

Photometric noise is only one of several sources of random noise in the measurement of radial velocities, and sources of systematic error must also be considered. However, in the context of this Workshop on accurate photometry, this paper will only present our experience with the accuracy with which we can measure the relative intensities of the various orders of interferometric transmission through the Fabry-Perot etalon, on a time scale of a few hours. Note that this differential photometry is done with a much more favorable optical configuration than our direct imaging described above. The images whose relative intensities are to be measured (Figure 7) are extremely stable in shape, size, and position because they are images of a local object (the output end of the fiber waveguide), instead of stars seen through miles of atmosphere. Furthermore, we are measuring the relative brightness of one star at different wavelengths instead of the relative brightnesses of different stars. Finally, the monochromatic images of the fiber optic output cover many more pixels (see Fig. 7) than our images of stars, thus reducing the importance of spatial nonuniformity of the array of detecting pixels. Perhaps not surprisingly then, we achieve a signal to noise ratio of several hundred with our short-term differential photometry of the Fabry-Perot interference orders. More specifically, the standard deviation of repeated measurements of the ratio of a pair of well-exposed Fabry-Perot orders is about $\pm$ 0.14 %, and we have not detected any deviations of the signal from linearity at this level of accuracy.

III.C.5. Results of Observations

Is this level of accuracy sufficient to measure changes in the radial velocities of stars to +_ 5 meters per second per year? A discussion of all the sources of random and systematic errors in measuring radial acceleration is beyond the scope of this paper; here we address only our requirements on random photometric errors. Even if a star is observed many times in a year, a single observation still should be accurate to +_ 15 meters per second. To achieve this, assume that the contribution to this error from photometric noise must be kept less than +_ 3 meters per second per observation. In our spectral region (4000 to 4500 Angstroms) the average absolute value of the slope of the solar spectrum is about 2.3 times the continuum intensity per Angstrom. For example, a Doppler shift of 0.01 A will introduce, on the average, a change of +_ 2.3 % in the intensities of the spectrum points. A Doppler shift of +_ 3 meters per second corresponds to an average photometric change of +_ 0.0097 % in the spectrum points of this wavelength region. If the observations of N independent spectrum points (interferometric transmission maxima) are used to detect acceleration along the line of sight, then the intensity at each point must be measured with a photometric accuracy of +_ 0.0097*SQRT(N) per cent. It follows that a large number N of independent spectrum points must be observed; the large format of the RCA CCD and our cross-dispersed spectrometer optics give us that large number, as can be seen in Figure 4. Currently our short-term random errors in differential radial velocity range between +_ 6 and +_ 20 meters per second per observation, depending on the brightness of the star, its spectral type, and the current calibration errors. See Figure 8.

IV. Suggestions for a Multi-Star Differential Photometer

Recently there has been interest in a new type of photometer that could observe simultaneously several stars separated by tens of arcminutes in the sky. Because panoramic detectors such as CCD's have a large number of independent resolution elements it is reasonable to ask whether such a detector could be adapted to the problem of monitoring several bright stars distributed over an area in the focal plane that is much larger than the CCD itself. Here we discuss one of the methods that has been proposed to do this.

IV.A. Fiber optic feeds from stars to CCD

Feeding the light from each star to the CCD by means of individual optical fibers that are bundled at their output ends has several advantages. There would be no limit on the angular field over which the stars could be distributed because the fibers would be individually jacketed and mounted, and could have any length. Microlens optics on the input end of each fiber could image the telescope objective, instead of the star, onto each fiber so that changes in the position, shape, and size of the star image would have little effect on the amount of light entering each fiber. Furthermore, changes in the spatial distribution of illumination on the telescope objective would not be preserved at the output end of the fibers because of the way fibers "scramble" the distributions of directions and brightnesses of the incident rays. Thus the output end of each fiber would always have the same spatial distribution of brightness. The output ends of the fibers, once brought together, could then be imaged onto the CCD. Photometry would be much more accurate than in the

case of direct imaging because the sizes, shapes, positions, and distributions of brightness of the objects being measured would be extremely stable on the CCD. Unless the stars are bright, some of the individual fibers would have to be dedicated to monitoring sky background. In the case of differential photometry a map of the sensitivity of the CCD ("flat field") would not be necessary, thus eliminating another source of noise.

IV.B. Re-image fiber outputs onto detector

We recommend that the bundle of fiber ends be re-imaged on the CCD by a lens instead of contacting them directly on the CCD. A relay lens avoids physical damage to the fragile CCD chip, allows it to be recessed well inside a refrigerated chamber while the optics are outside at ambient temperature, and allows different detector packages to be swapped under the same optics. The images of the fiber ends should be well separated on the CCD to avoid contamination. The magnification factor of the relay lens should be chosen so that the images of the fiber ends have diameters of 10 or more CCD pixels.

IV.C. Readout system

The signal from each CCD pixel should be digitized to 14 bits after a slow, low-noise readout. Full scale on the A/D should be slightly higher than the saturation level of the CCD, to obtain the greatest possible dynamic range of signal. On some types of CCD's, including the RCA SID 53612 and SID 501EX, the signal from several adjacent pixels can be integrated (analog-summed) before readout through the on-chip preamplifier; this procedure is recommended in cases such as this in which only the total brightness of a large sub-array of pixels is desired, because the readout noise, which originates in the on-chip preamplifier, is only applied once to the integrated signal instead of to each pixel individually. This approach could reduce the effect of electronic noise by a very large factor.

We are pleased to acknowledge the assistance of N. Castillo, K. Denomy, L. Doose, R. Goff, S. Marinus, G. McLaughlin, and T. Mullin. This work was supported by Gehrels' NASA contract NAGW-3454, McMillan's NASA grant NAG2-52, and McMillan's NSF grant AST-8117261.

V. References

Fowler, A., Waddell, P., and Mortara, L. 1961, "Evaluation of the RCA 512x320 CCD Imagers for Astronomical Use", in Proc. S. P. I. E. vol. 290, "Solid State Imagers for Astronomy", pp.34-44.

Geary, J. C., and Kent, S. M. 1980, "Imaging characteristics of the RCA 512 x 320 CCD", in Proc. S. P. I. E. vol. 290, "Solid-State Imagers for Astronomy".

McMillan, R. S., and Stoll, C. P. 1982, "Software Simulations of the detection of rapidly moving asteroids by a charge-coupled device" in Proc. S. P. I. E. vol. 331, "Instrumentation in Astronomy IV", pp. 104-112.

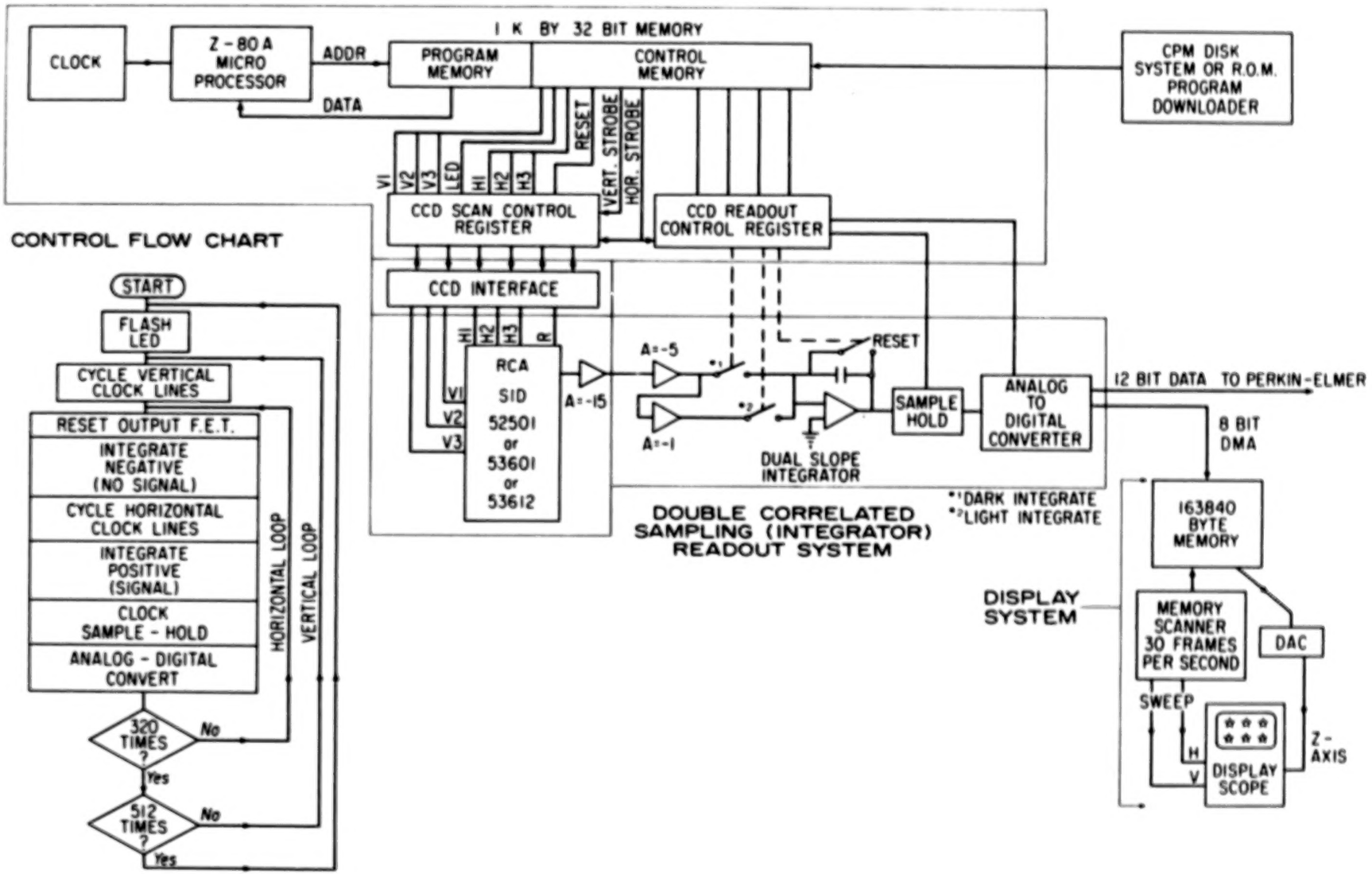


Figure 1. CCD control and readout electronics.

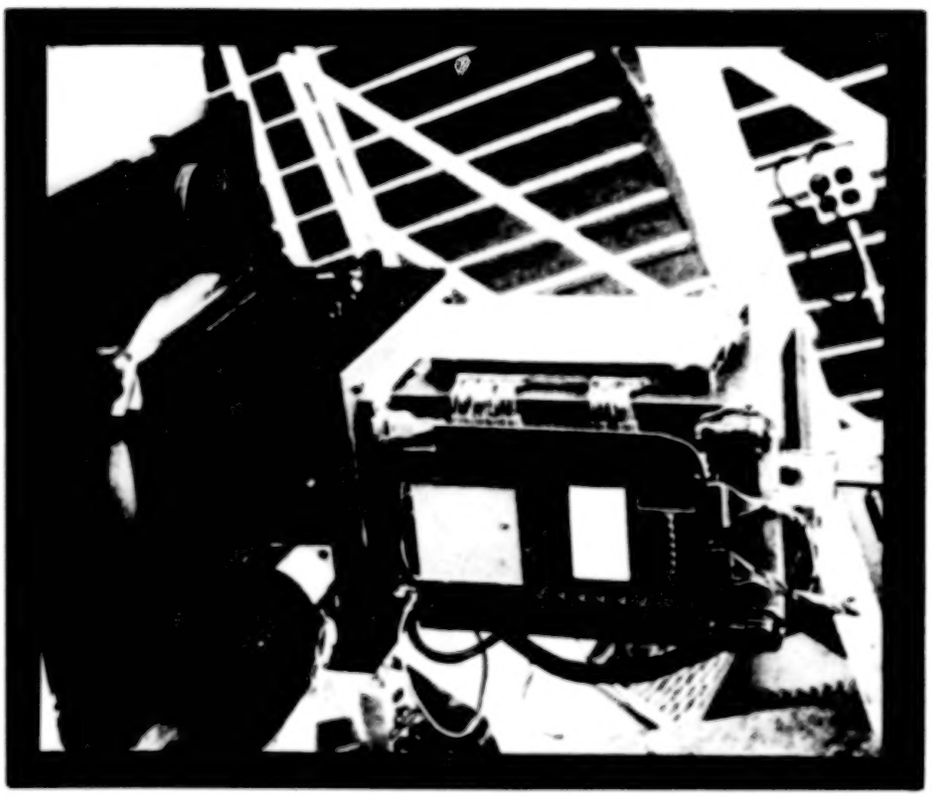


Figure 2. The CCD refrigeration chamber and Z80 control electronics mounted at the Newtonian focus of the University of Arizona Observatories' 0.91-meter telescope.

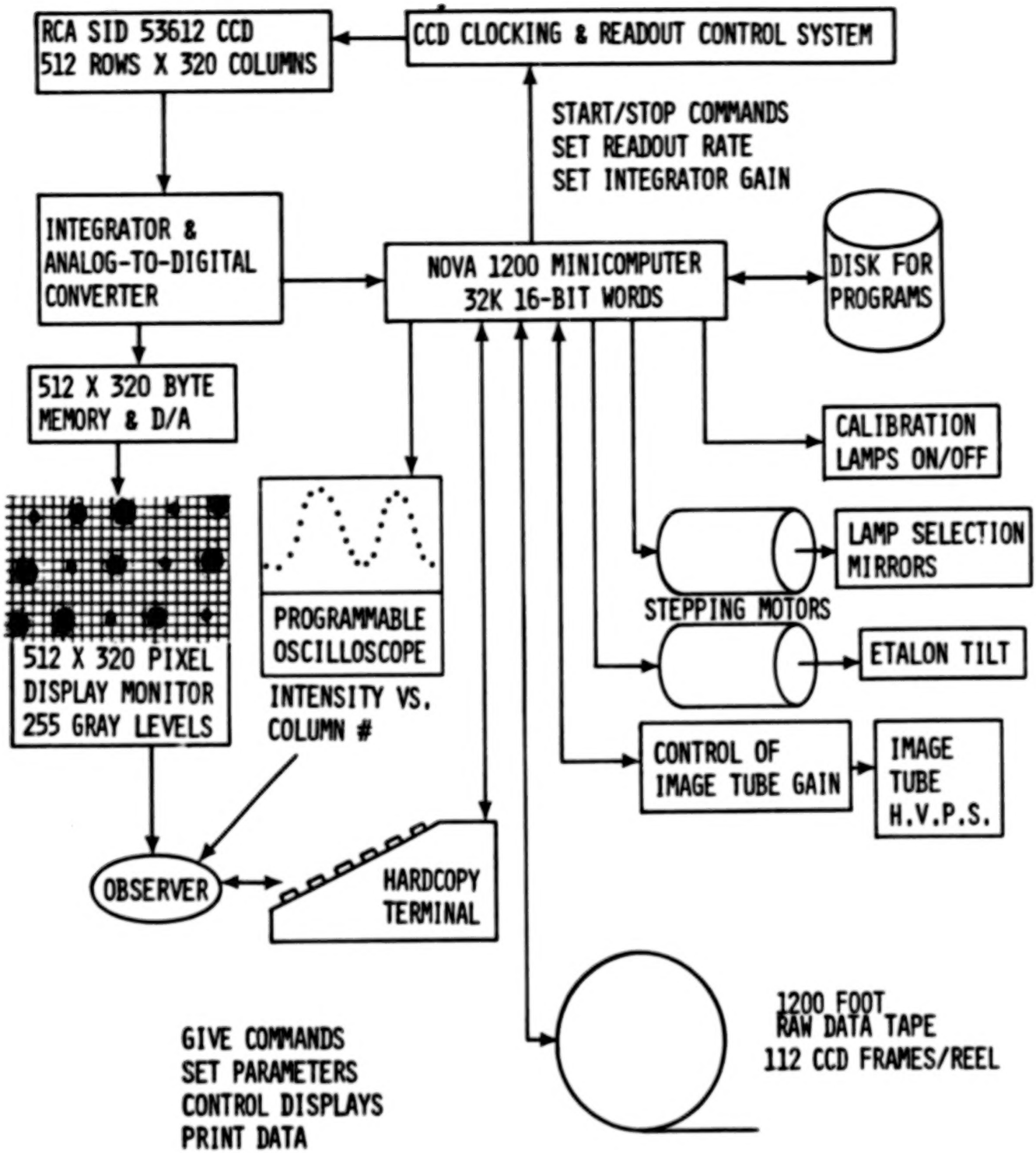


Figure 3. Schematic of the real-time control and display system.

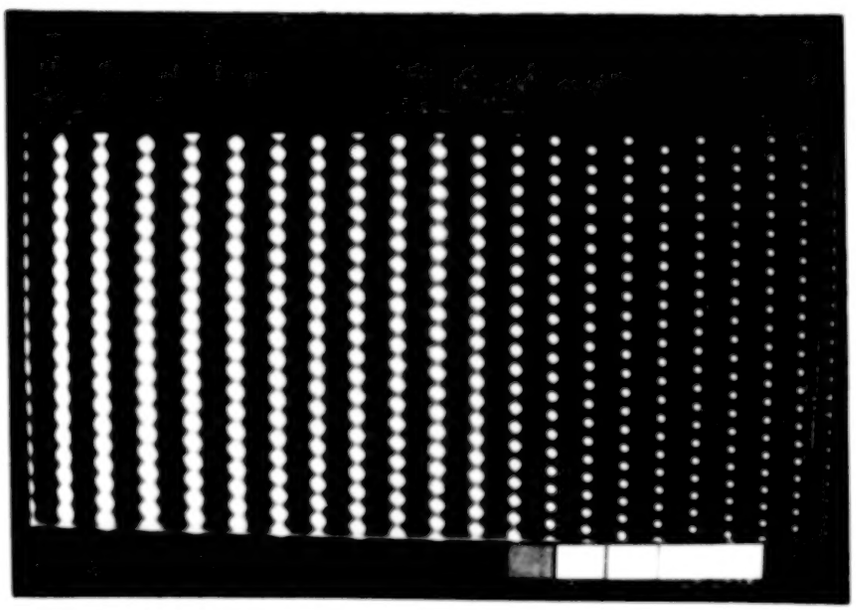


Figure 4. White lamp spectrum on the RCA SID 53612 320x512 CCD detector array. The echelle dispersion runs from top to bottom; the cross dispersion from side to side.

Figure 5. A section of a CCD drift scan near the galactic pole.

LPL STELLAR ACCELEROMETER: JUNE 1983

OPTICAL SCHEMATIC, NOT TO SCALE

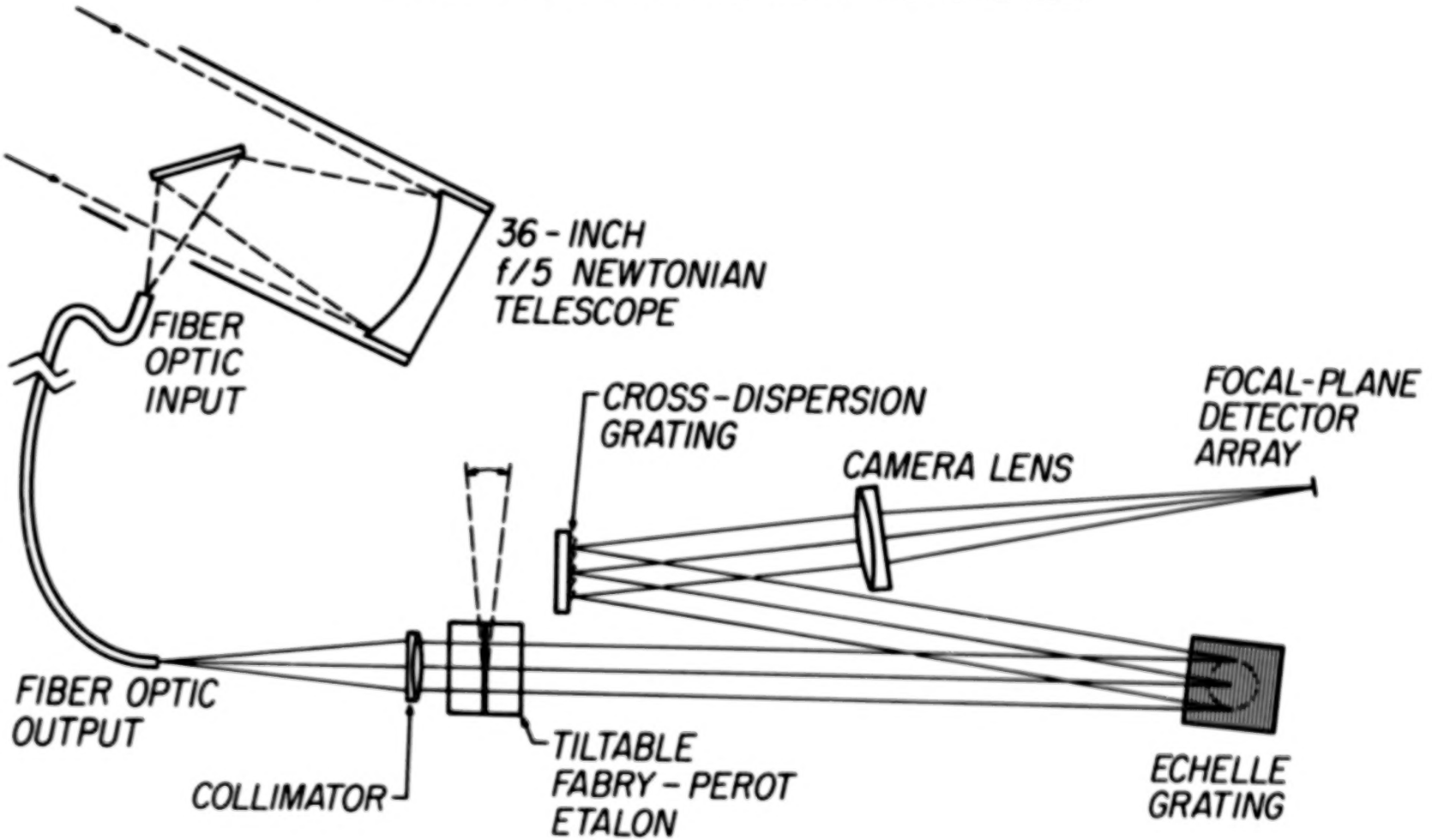


Figure 6. Schematic of the radial accelerometer.

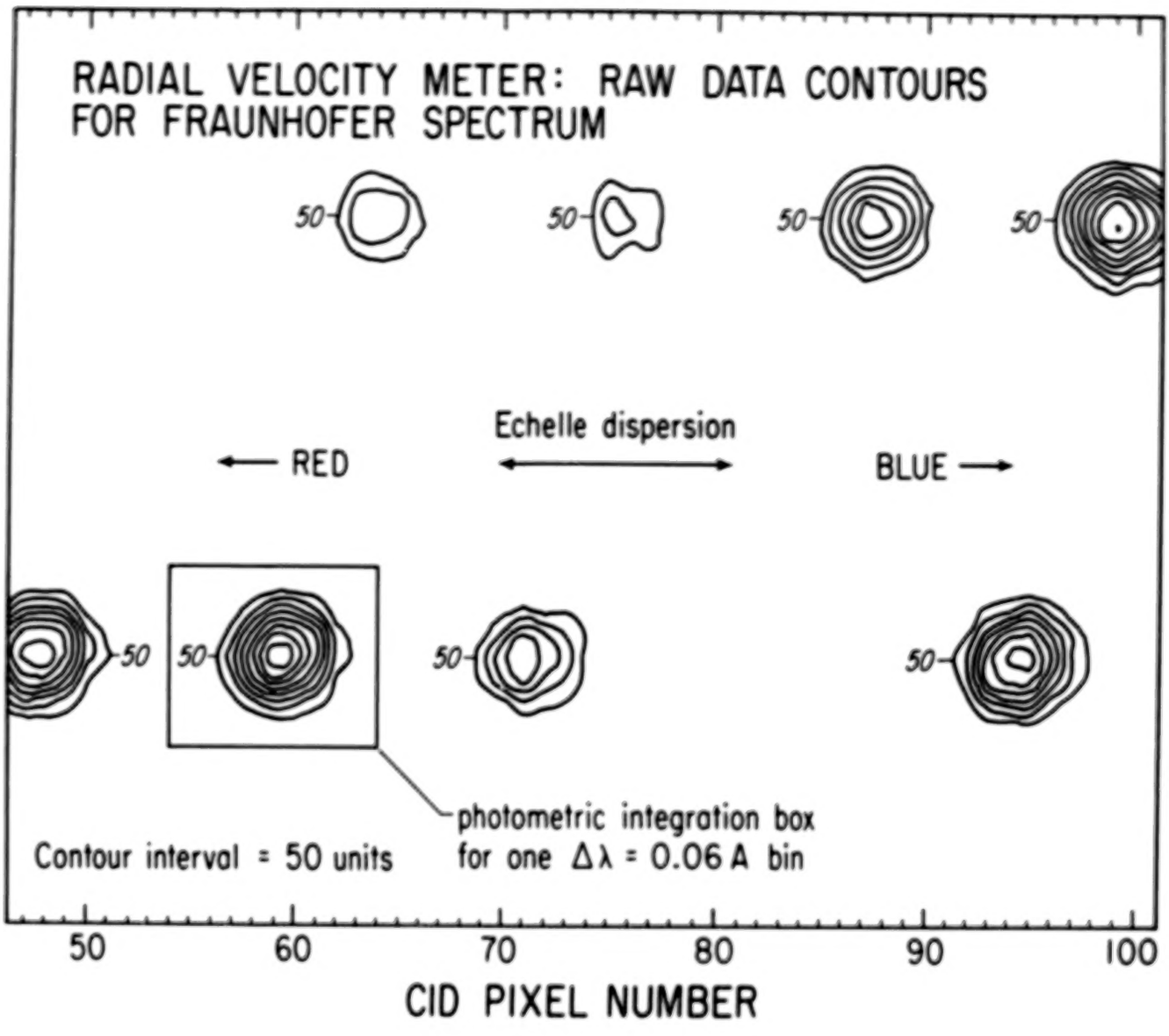


Figure 7. Intensity contour map of part of an image of the solar spectrum on the detector array. Each "hill" is a peak of transmission through the Fabry-Perot etalon.

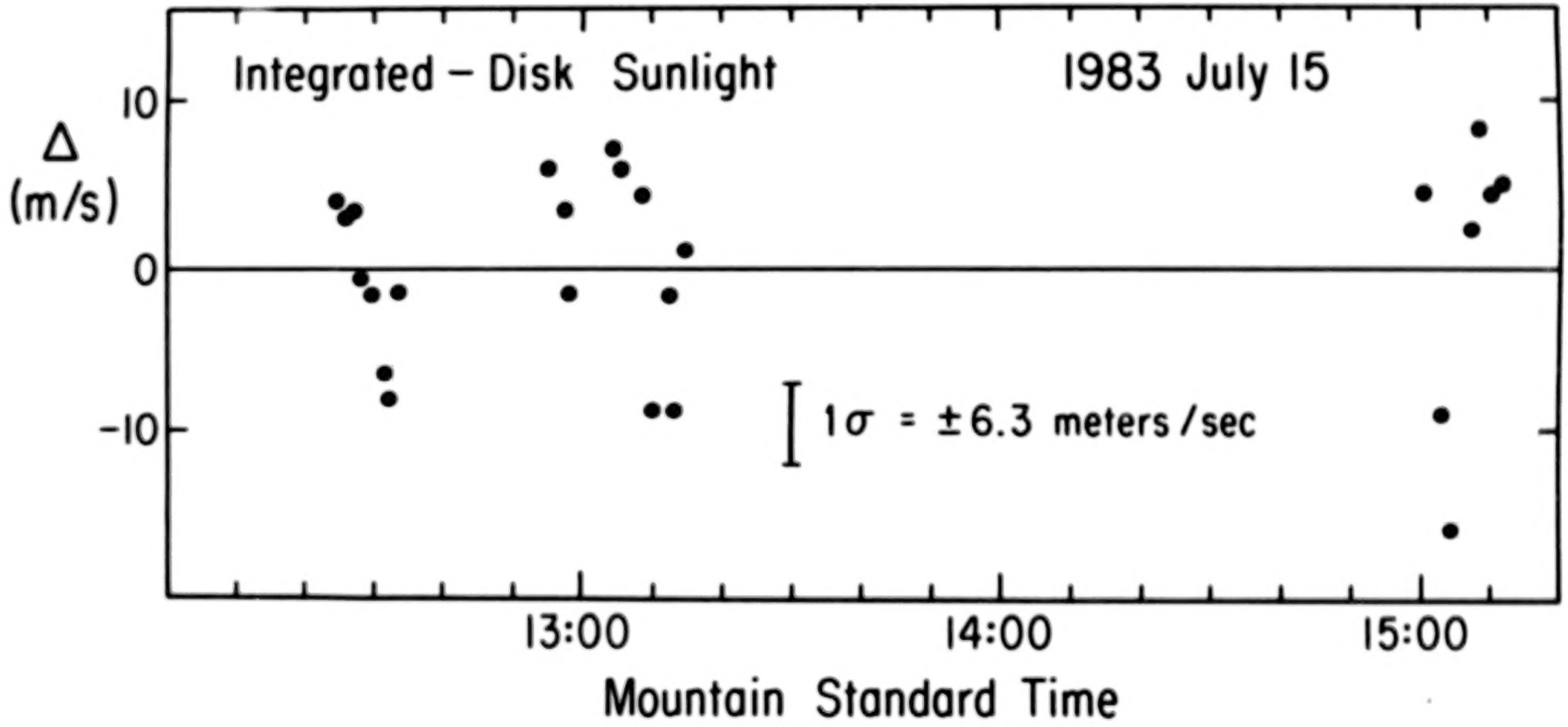


Figure 8. Residuals of differential radial velocity measurements of light from the integrated solar disk, compared with the predicted variation of velocity along the line of sight due to the rotation of the earth. The theoretical curve has been subtracted from the data to show the residuals more clearly.

D13

CCDs : Their Cause and Cure

S. Djorgovski
Astronomy Department
University of California
Berkeley, CA 94720

1. Introduction

Charge coupled devices (CCDs) have opened new horizons in the optical astronomy. We can now do photometry down to ~26^m and spectroscopy down to ~23^m. Such limits were simply unthinkable with the detectors of the previous generation. Practically every current issue of the major journals bears witness to these advances.

However, CCDs are still something of a novelty. We are now in the "1P21 era" of CCD technology, and we can expect many improvements in the years to come. Another caveat is that most of the presently existing astronomical CCD systems are oriented and designed primarily for work on faint objects. The use of CCDs for high-precision work on bright objects remains to be explored. As any other type of detectors, CCDs have their own specific problems and limitations. Moreover, most users are still far from extracting from their CCD data everything that they can get.

This paper is an attempt to discuss some of those issues and problems, and some methods of coping with them. Only the problems specific to CCDs are covered here, though some remarks are also applicable to other digital imaging detectors. Despite the title, this paper does not aspire to the level of completeness and depth of the classical paper on photomultipliers by Young (1974). All the lore about the atmospheric, telescopic, etc., problems and limits to good photometry still applies. Those and other related issues are discussed in the other papers in this volume. This paper does not discuss any engineering aspects, nor CCD photometer design. It is written from the viewpoint of a CCD user, and it is refracted through a foggy prism of the author's own interests, bias, ignorance, etc.

Evaluations of CCDs for astronomical use, and some applications, can be found in the papers by Leach et al. (1980), Gursky et al. (1980), Janesick, Hynecek & Blouke (1981), Fowler, Waddell & Mortara (1981), Geary & Kent (1981), Zucchino et al. (1981), Mortara & Fowler (1981), Robinson (1981), Goad & Ball (1981), Lauer et al. (1983) and McGuire (1983). Wright's thesis (1982) contains both engineering detail and useful application information. Many interesting papers can be found in SPIE vols. 264, 290, 331, and 445.

In learning the CCD trade and preparing this paper, I have greatly benefited from conversations and troubleshooting sessions with many of my colleagues at Berkeley, Lick, and Kitt Peak. They are too numerous to mention individually, and I thank them all. Support from the NSF grants AST81-18557 and AST81-16125, and the NASA contract NAS5-28086 is also gratefully acknowledged.

2. Photometry with CCDs

The reason why we can do so well with CCDs is threefold: first, they are high--quantum-efficiency (QE --> 1) detectors, very linear, with high dynamical range; second, they are imaging detectors, which enables one to use much of the spatial information and perform local simultaneous background subtraction; and third, they are Poissonian, white-noise detectors, friendly for digital image stacking and image processing.

The imaging character of CCDs makes them particularly well suited for photometry. The full spatial information enables astronomer to to input any *a priori* knowledge he may have about the measured object, and thus reduce the number of degrees of freedom in the measurement.

For example, in the case of surface photometry one can assume (in most cases, based on image appearance) that the isophotes are well represented with concentric ellipses. One can then fit ellipses to the data, and determine runs of surface brightness, ellipticity, position angle and parametrized deviations from elliptical shape, as functions of the major semiaxis. If *a priori* knowledge were not used, one would have to measure many more radial light profiles, all of which would be noisier and far more difficult to interpret. The methods of surface photometry on a digital array are discussed by Williams & Schwarzschild (1978), Kent (1983), Cawson (1983), and Lauer (1983).

In the case of stellar photometry, one can predetermine the point-spread function (PSF) from the data, and fit stellar images to it. This produces four numbers: accurate center coordinates, background value (or parameters of the surface fitted to the background), and the PSF scaling factor, which conveys the magnitude. This procedure avoids the pitfalls of close companions, nonsimultaneous/nonlocal background determination, and asymmetric centering which are inherent to aperture integrating measurements. Not only is spatial information provided, but one can also devise a fit with appropriate statistical (Poissonian) weighting: the pixels with most signal should get higher weight in the measurement. This technique is particularly useful in crowded fields, where it can deblend close stellar images. The theory behind this method is given by King (1983). Practical applications are described by Tody (1980 and 1981) and Newell (1983). It should be noted that a good spatial sampling is required: if the pixels are too large with respect to the PSF width, some spatial information is lost; and if there is some intra-pixel response variation, the results may depend on the PSF center position within the pixel.

The advantages of imaging detectors (e.g., CCDs) over aperture integrating detectors (e.g., photomultipliers) are summarized in the Table 1. This comparison does not include any possible quantum-efficiency differences.

Good CCD stellar photometry achieves night-to-night repeatibility of the order of $5 \cdot 10^{-3}$ at a good site, in circular pseudo-aperture measurements; PSF fitting with the uniform weighting narrows the residuals by some 25% (Seitzer, pers. comm.) ; a Poisson-weighted PSF fitting should do better yet.

A discussion of some methods, problems and solutions in faint object CCD spectroscopy is given by Djorgovski & Spinrad (1983).

TABLE 1

Advantages of imaging detectors for stellar photometry

APERTURE INTEGRATING DETECTOR (e.g., photomultiplier, silicon diodes, etc.)	IMAGING DIGITAL DETECTOR (e.g., CCD, SIT vidicon, microchannel-plate detector, etc.)
Cannot have local and simultaneous background measurement.	Local and simultaneous background easy to determine.
Detector surface response nonuniformity is hard to calibrate.	Response nonuniformity information is easy to obtain; flatfielding is a standard procedure.
Uniform weighting over over the whole aperture.	Poisson weighting easy to implement.
Glitches (cosmic rays, etc.) are hard to remove.	Glitches are easily isolated and removed from the measurement.
No spatial information available (e.g., close companions, nebulosities).	Full spatial information; source is easy to isolate.
One object measured at a time (different atmospheric conditions for different objects).	Many objects measured at the same time (same atmospheric conditions for all objects in the field).
Object centering in the aperture is often subjective, and may cause systematic errors.	Impersonal object centering is a standard procedure.
A priori knowledge about the object structure cannot be used.	A priori knowledge about the object can be used profitably (e.g, PSF or isophote fitting).
Seeing excursions make small apertures hard to use accurately.	Resolution is limited by pixel size.

3. Physical Properties of CCDs

The working principles of CCDs are described on an introductory level in Amelio (1974), Kristian & Blouke (1982), and on a more technical level in Blouke et al. (1981) and Wright (1982). The work of Wright (1982) is particularly rich in various technical aspects, and describes the different types of CCDs. A wealth of technical and engineering details can be found in the notes privately distributed by J. Janesick of JPL.

The important structural characteristics of CCDs are that they are integrating, "self-scanned," photon counting (one electron for each detected photon), solid state devices. The typical physical sizes are of the order of 1 - 3 cm, and typical raster formats of 500x500 pixels (TI), 320x512 (RCA), 385x576 (GEC), 800x800 (TI), and 1500 x1500 (GEC) should be available soon. Typical (and optimal?) operating temperatures are around -100° C, with liquid N_2 as the most common coolant. Some CCDs are physically warped. This causes focus variations across the surface, which amount to a spatially variable PSF.

Other relevant properties of CCDs as detectors and a comparison with some other astronomical detectors are given in Table 2. Figure 1 shows a typical spectral response curve.

Most "untreated" CCDs (except for the RCA thinned chips) have very poor blue response. Two ways have been devised to help this. First, one can cover the CCD with a layer of the organic phosphor Coronene, whose luminescence converts the incoming blue photons into longer, detectable wavelengths. In the second method, a certain amount of electric charge is deposited on the back of a CCD. This apparently causes changes in the electronic energy-band structure of the silicon crystal used for the detection, which manifest themselves both by a dramatic increase in the blue sensitivity, and improvements in the response uniformity across the chip. This "backcharging" is now commonly accomplished by flooding the device with UV light before it is cooled to the operating temperature. The infrared response of a CCD can be extended somewhat by operating it on a higher temperature; however, this scheme still needs some experimentation and development.

CCD readout is based on charge transfer. A pocket of charges acumulated in any pixel is clocked through hundreds of pixels before it is read out. This process is imperfect: a part of the pixel's charge will be left in the following pixel(s). L. Goad's experiments (pers. comm.) with a TI 800x800 chip show that this can be modeled as a linear process:

$$\{\text{charge left in the trailing pixel}\} = A \cdot \{\text{charge in the leading pixel}\} + B$$

where $A < 10^{-2}$ and $B \approx 100 - 400$ electrons (after the complete readout).

The presence of a constant term B causes a signal-dependent PSF at the low light levels. Together with the low-end deferred charge nonlinearity (Sect. 8), this may suggest use of a CCD preflashing procedure for some applications. Diffusion of charge across the readout columns is very small, typically of the order of 10^{-5} to 10^{-4} of the total pixel signal. Note also that charge transfer efficiency increases with the temperature, so that CCDs should not be overcooled.

TABLE 2

Physical properties of CCDs and comparison with some other detectors

	Photographic plate (IIIa-J)	Photomultiplier (S20)	CCD
Maximum Q.E.	~ 2 %	~ 20 %	~ 80 %
Useful spectral range (Å)	~ 3000 - 5000 (I-Z emulsion: up to ~ 11000 Å)	~ 1200 - 7500 (S1 cathode: up to ~ 11000 Å)	~ 4000 - 11000 (with UV flood : down to ~ 2500 Å)
Dynamical range	~ 20 - 25 dB per pixel	high	~ 35 dB per pixel
Typical pixel size (aperture size)	10 - 50 μ	large (few mm ?)	15 - 30 μ
Useful detector size	up to ~ 50 cm	~ 1 cm	~ 1 - 3 cm
Linearity	poor	excellent (at low count rates)	good to ~ 10^{-3}
Surface response uniformity	~ 1 %	~ 2 - 5 %	raw: 1 - 100 % flatfielded: ≤ 1 % drift scan: ≤0.1 %
Repeatibility	poor (few %)	good	excellent
Principal noise source	emulsion grains; plate fog	Poissonian (photon count)	Poissonian (phot. count); readout (amplifier)

Note: SIT and ISIT vidicons closely match the CCDs in many properties; they tend to have even higher dynamical range, but lower Q.E., higher noise, and larger deviations from linearity. The same is also true for microchannel-plate detectors, but they have a better time resolution.

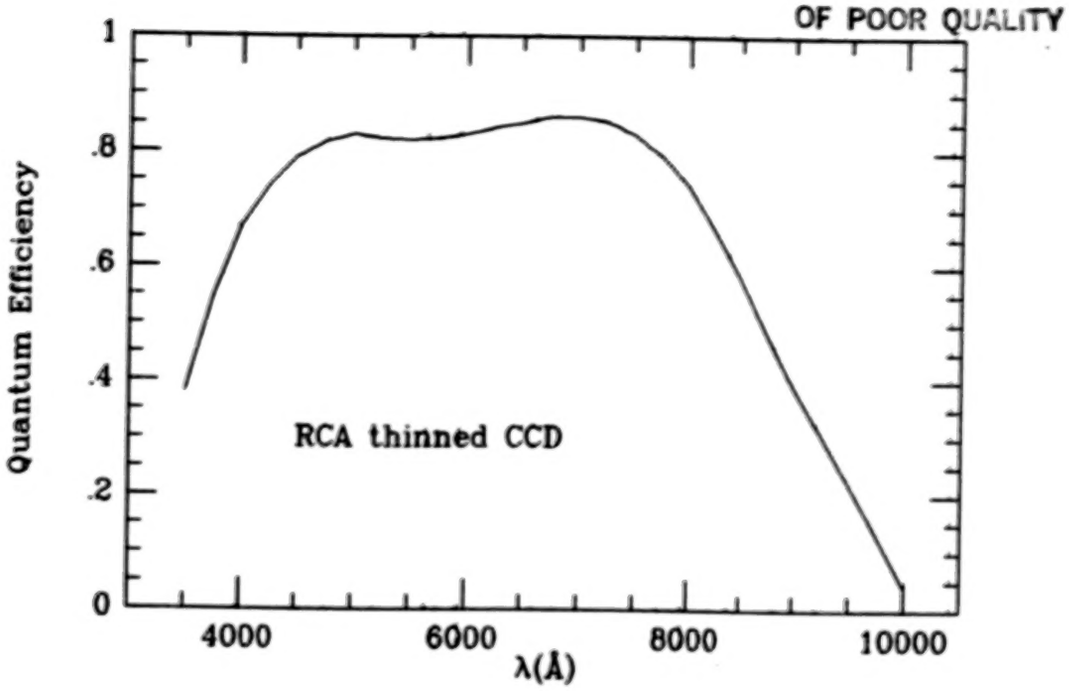


Figure 1. Response curve of a thinned RCA CCD (after Geary & Kent 1981).

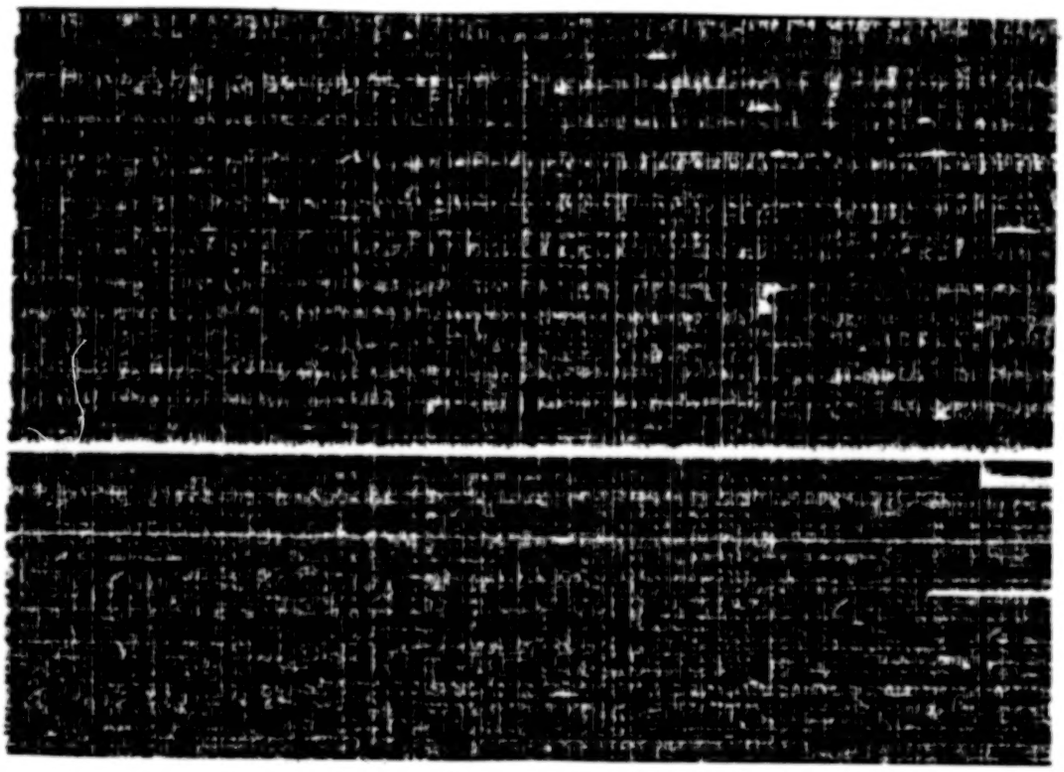


Figure 2. Highly enhanced low-level periodic "tweed" pattern in the Lick 576x385 GEC CCD. The amplitude is ~ 1 DN (readout unit). Poor-charge-transfer columns (horizontal white streaks) are also evident.

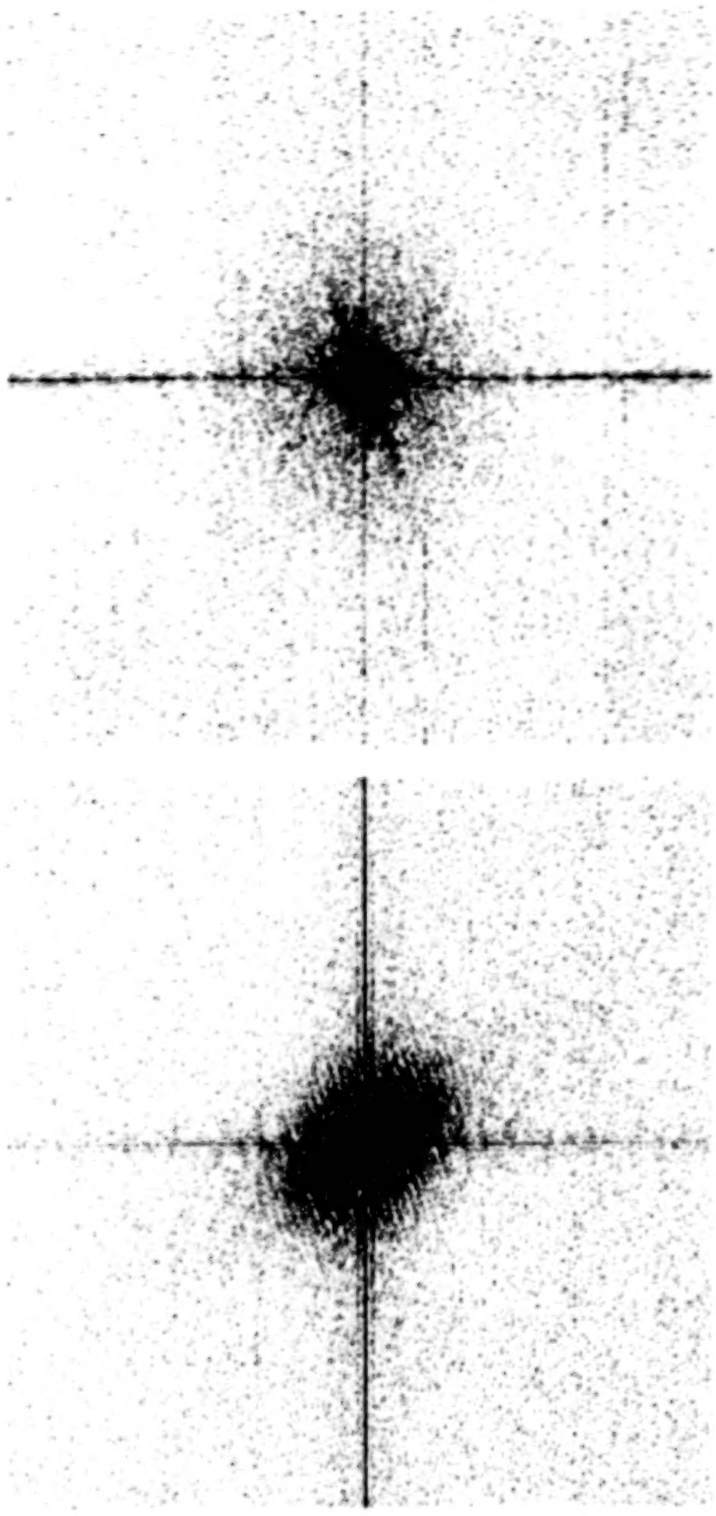


Figure 3. Two-dimensional log power spectra for 256^2 sections of the Lick TI 500x 500 CCD (top) and GEC 576x385 CCD (bottom) sky exposures. The centrally concentrated fuzz is from the data signal; excess power on the Fourier axes (k_x=0 and k_y=0) is due to column-to-column readout periodicities; diagonal fringes are due to various periodic artifacts.

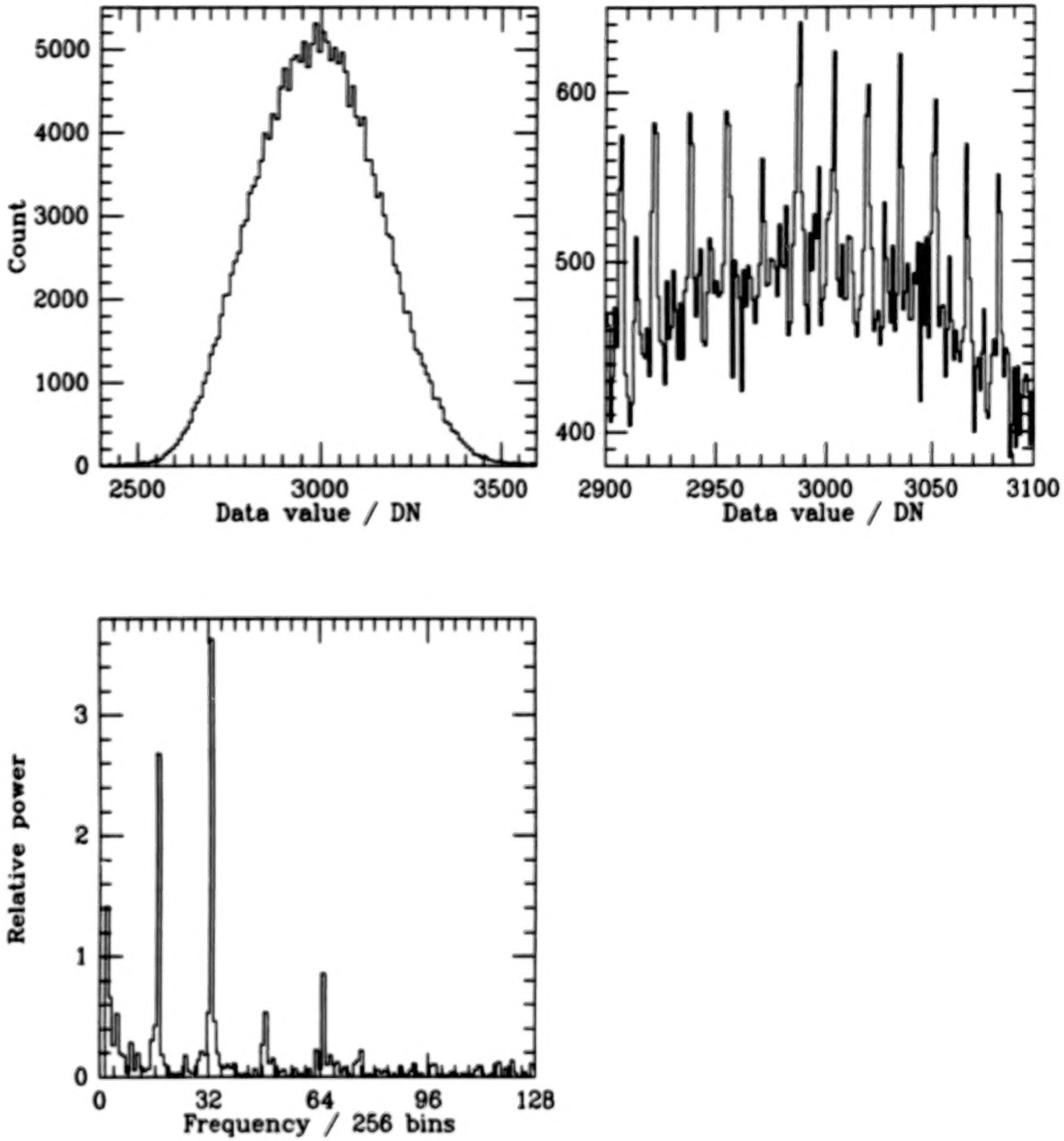


Figure 4. (Top left) Data histogram for a flatfield exposure with the Lick GEC CCD. (Top right) Zoom-in on the histogram peak with higher resolution reveals the 16-bin periodic ADC bias. (Bottom left) Power spectrum of the high-resolution histogram shows clear evidence for the ADC periodic bias.

4. Electronics Troubles

Most present astronomical CCD systems go to the telescope before they are completely debugged. As a consequence, the readout electronics often inserts low-level spurious signals, patterns and other artifacts in the data.

A common problem is a periodic pattern, due to multiple resonances of internal frequencies in the readout electronics and the 60 Hz pickup. Periodic patterns may also be due to the periodic instabilities in the electronics. An example of a high spatial frequency "herringbone-tweed" pattern is shown in Fig. 2. Lower-spatial-frequency patterns may be more common. The best way of diagnosing (and even fixing) this type of problem is through the use of 2-dimensional Fourier transforms, and filtering in the Fourier domain. Examples of a power spectra of such spurious signals are shown in Fig. 3. These patterns represent an additive error, and they are not stable in time, which makes them difficult to remove. Typical amplitudes are of the order of several electrons (or ~1 analog readout unit) per pixel. A discussion of similar problems is given in Lorre & Gillespie (1980).

The second type of electronics-caused problem is the binning bias in the analog-to-digital converter (ADC). This is easily diagnosed by plotting the data histograms and their 1-dimensional power spectra (Fig. 4). There is nothing that one can do about this, other than replacing the ADC.

These problems are sometimes the limiting factor in the accuracy achievable with present-day CCD systems. Fortunately, they are completely fixable and do not represent a fundamental limitation.

5. Linearity and Saturation Phenomena

The full-well pixel capacity for most CCDs is typically several tens of thousands of electrons. The "noise floor" (Sect. 6) ranges from several to several tens of electrons per pixel. Therefore, the dynamical range is typically ~ 35 dB / pixel in a single exposure. For most of this range, the linearity is excellent. Typical deviations from the linearity are in the range of $(1-5)\cdot 10^{-3}$ (Janesick, Hynecek & Blouke 1981; Fowler, Waddell & Mortara 1981). This is one of the fundamental limits to accurate photometry with a CCD. Some devices become notably nonlinear slightly before saturation.

The full-well maximum of few times 10^4 electrons suggests that the Poisson limit for a well-sampled star is in the neighborhood of 10^{-3}, or 10^{-4} at the very best, in a single exposure.

When a pixel is saturated, the excess charge "bleeds" along the readout columns and sometimes even across them. This in itself is not too damaging (the excess charge is easily "scrubbed" away with a few repeated readouts). The problem, however, is that some of the excess charge may diffuse into the chip substrate, from which it diffuses back during the following exposures. This creates spurious "ghost images", which can persist for a couple of hours. This trouble can be prevented by maintaining the chip substrate at a higher potential. Examples of saturation phenomena are shown in Fig. 5.

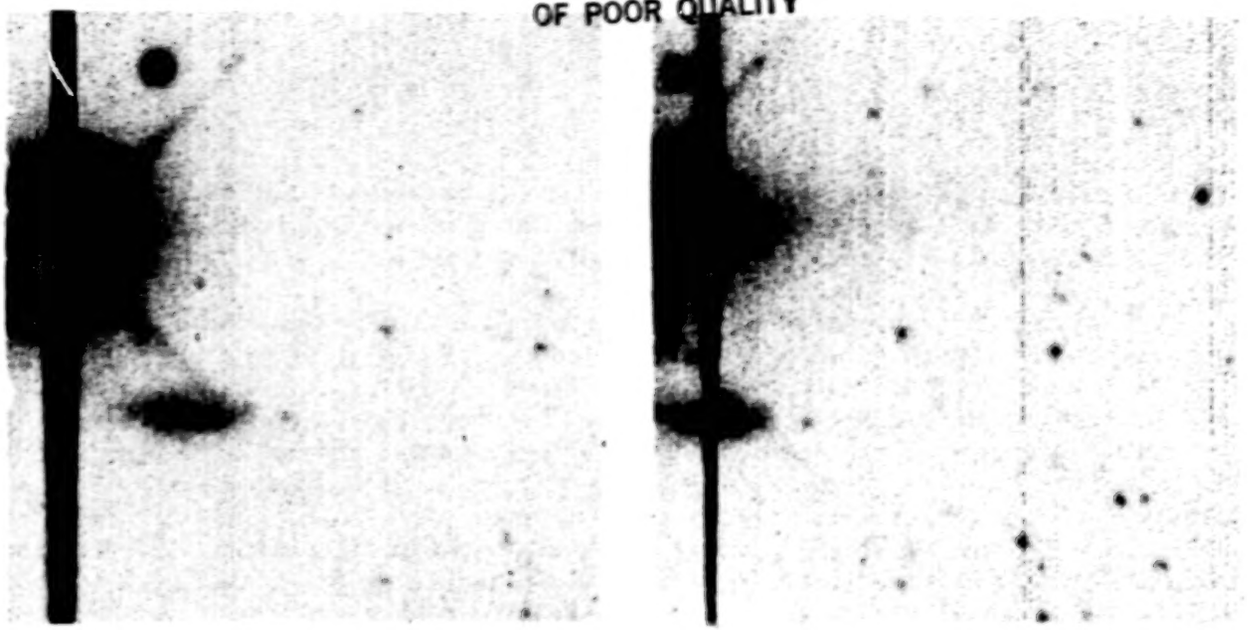


Figure 5. Examples of saturation phenomena. (Left) A very bright star leaves a charge "bleed" trail along the readout columns. (Right) In the subsequent exposure, the star has been moved to the left, but the charge diffusion "ghost" image is still present. The exposures were made with the Kitt Peak TI 800x800 PF CCD.

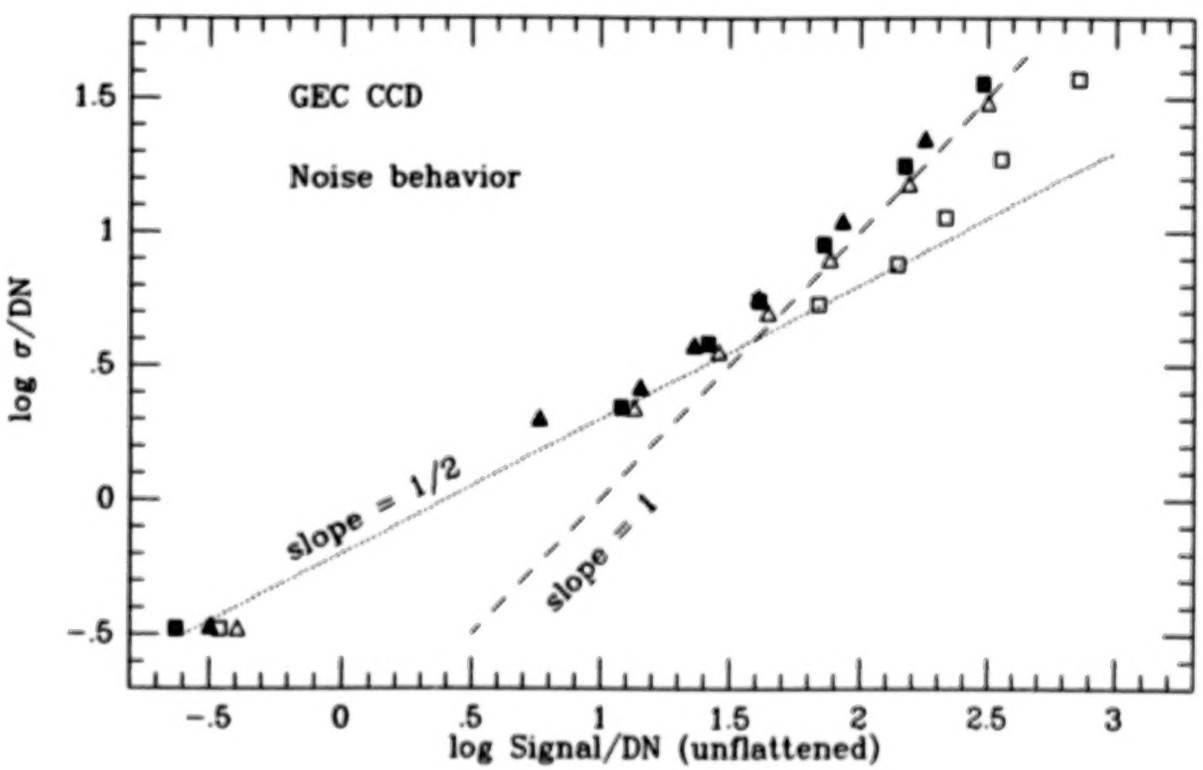


Figure 6. Noise regimes of the Lick GEC CCD. Log of the pixel RMS noise for a small subraster is plotted vs. the log of the mean signal. Different symbols correspond to different areas on the CCD. Explanation is given in the text (Sect. 7) .

6. The Noise Properties

The reason for cooling CCDs is to diminish the dark curent. It decays with the absolute temperature T as :

$$I_{dark} \sim \exp(\text{const.}/T)$$

At the usual operating temperature levels ($\sim 160° - 180°$ K) the dark current amounts to only a few electrons / pixel / hour, and so it can be neglected in most applications. For CCD chips with strongly non-uniform sensitivity, the dark current has a definite spatial pattern (Baum, Thomsen & Kreidl 1981).

The most serious source of noise in a CCD is readout noise, generated primarly by the on-chip amplifier. Presently, it ranges from ~ 5 electrons / pixel (for the best GEC devices), up to about 80 (RCA devices). This noise constitutes the "noise floor" of a CCD, and it is the main reason why CCDs win only in the long exposure and/or deep-imaging applications, rather than in the short-exposure ones. When presented as a readout time sequence, this noise has a "pink" power spectrum (Wright 1982).

Poissonian noise is inherent in the nature of the photon-count signal, and it is usually the major component of the net total noise.

For unflattened exposures, the sensitivity variations across the CCD constitute the pattern noise. It is a multiplicative error, and thus it is directly proportional to the signal. Proper flatfielding should remove this noise entirely.

The noise regimes of a CCD are depicted in Fig. 6. Similar figures are also presented in Janesick, Hynecek & Blouke (1981) and Fowler, Waddell & Mortara (1981).

7. Flatfielding

The sensitivity (response) of a CCD is not uniform across the surface. Non-uniformities range from a few percent for most modern devices, up to 200 - 300 % for some of the older ones. This multiplicative error is in principle completely removable by division with a picture of a uniformly illuminated surface (commonly refered to as a flatfield), in the process known as flatfielding. It is possible to do this with a very high accuracy. A particularly dramatic example is shown here in Figs. 7a,b.

It is useful to distinguish between the global (low spatial frequency), pixel-to-pixel (high spatial frequency) and intra-pixel (super-Nyquist frequency) sensitivity variations.

The only measurement of intra-pixel sensitivity variations known to the author is given in Wright (1982), for a GEC device. The response varies in a wavy fashion, parallel to the electrodes, with 10 - 20 % amplitude. It is probably caused by the electric field distribution within the pixel, but there also seems to be a slight wavelength dependence, indicative of possible interference effects.

Figure 7a. A 465x500 section of a Lick TI 500x500 CCD raw exposure of NGC 4382. Sensitivity varies by ~ 200 - 300 % across the chip, with very sharp drops at the edges of the "bird" pattern. This is a logarithmic picture with 1^m = 4 dB per gray scale cycle.

Figure 7b. Exposure from Fig. 7a, flattened with a sky median frame, in the same logarithmic representation. The "bird" pattern is practically gone, showing only as a higher noise area. The bad columns have been masked.

Most unfortunately, a CCD sensitivity pattern is wavelength-dependent. This is easily demonstarted by plotting the histograms of flatfields illuminated with light of different spectral composition, as demonstrated here in Fig. 8, and even better by Wright (1982). This is a severe limit on the effectivness of flatfielding. Note, however, that this difficulty does not occur in CCD spectroscopy, since the flat-field exposure is dispersed in the same way as the sky exposures. Similarly, the trouble is much smaller in narrow-bandpass applications.

The common ways of obtaining the flatfields are:

- Dawn or twilight sky exposures. These give a good signal, but they are far too blue to be useful in most applications.
- Incandescent lamp projected on a uniform white surface out of focus. This also provides a good signal (note that many such exposures can be co-added), but again it is a poor match to the sky color. Moreover, in most cases, the global illumination pattern of the lamp flatfield is not the same as that of the sky, due to focus differences and possible specular reflection from the illuminated white surface. An example of this is shown in Fig. 9.
- Sky median frames. They have the perfect spectal composition, and convey the "true" low spatial frequency variations, but tend to have low signal, and thus poor high-spatial-frequency characteristics. Sky median frames can be produced as follows:

A number N (at least ~ 5) of exposures free of bright and/or very extended objects are obtained during the night, and rescaled to a common median data value. It helps to run data through a preliminary object detection, and flag out all the pixels suspected of belonging to objects (stars, galaxies, etc.) . Let I_{ijk} be the intensity of (i,j) pixel of the k-th rescaled input frame. The output sky median pixel O_{ij} is computed as a mode estimate of the k-sequences of all non-flagged I_{ijk}. The mode estimate is in practice evaluated either as the median of Iijk (k varying), or the weighted average of the median and its adjacent values. This non-parametric statistic discriminates against the object pixels, situated in the high tail of the sequence. In this way, a virtually object-free sky frame is produced. But not quite - if N is small, there is always some statistical leakage of faint objects or faint envelopes of bright objects into the sky median frame.

If a median-for-the-night frame is obtained in this fashion, it has a good color and good low spatial frequency response, but typically poor signal and possibly some faint residuals of objects from the input frames. A median-for-the-run (or for a longer period) produced in the same way is free from these defficiencies, but it may have different low spa-tial frequency response.

The next step is to combine the low-frequency information from the median-for--the-night with the good pixel-to-pixel statistic (high-frequency information) from the median-for-the-run. This is accoplished as follows:

1) Divide the low frequency frame by the high frequency frame.
2) Partition the resulting frame into a grid of subframes (cells), e.g., 20x20 pixels or larger, and evaluate the mode estimator in each cell.
3) Fit a low-order polynomial surface to the grid. A bicubic fit is usually sufficient.
4) Multiply the high-frequency frame with this surface. This is now the final flatfield frame.

If the data-frame signal is dominated by the counts, this type of flafield removes the nonuniformities down to the Poisson limit. If there are bright objects in

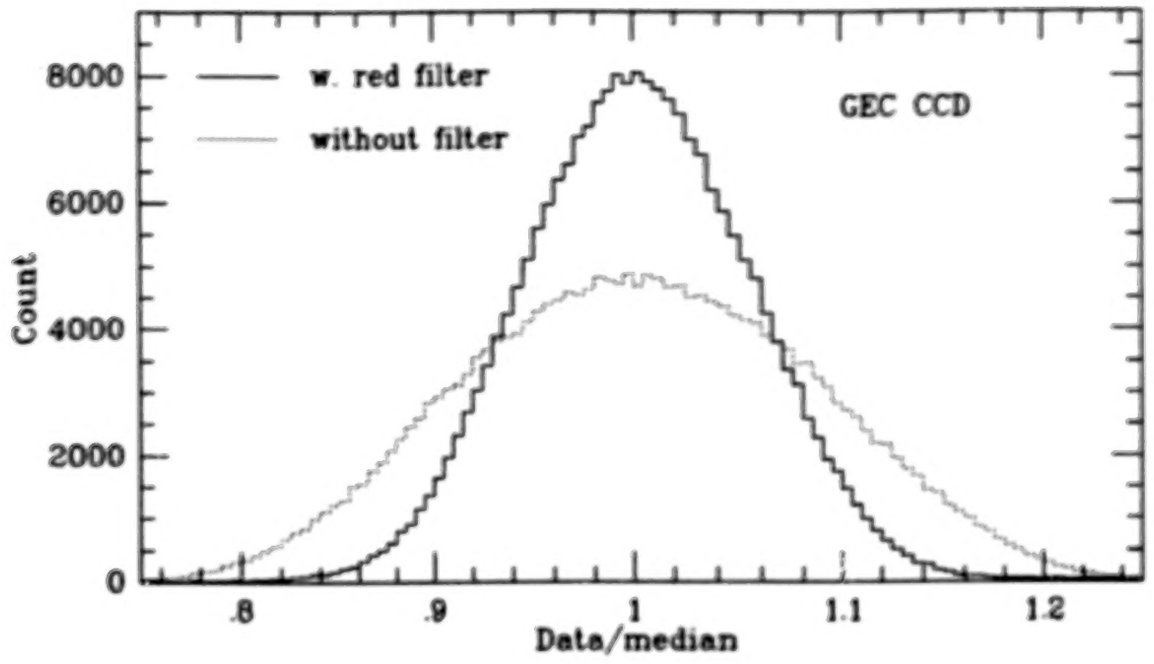


Figure 8. Wavelength-dependent response of the Lick GEC CCD demonstrated as a difference of relative data histograms for flatfields illuminated with light of different spectral composition. The response is more uniform in the narrower bandpass.

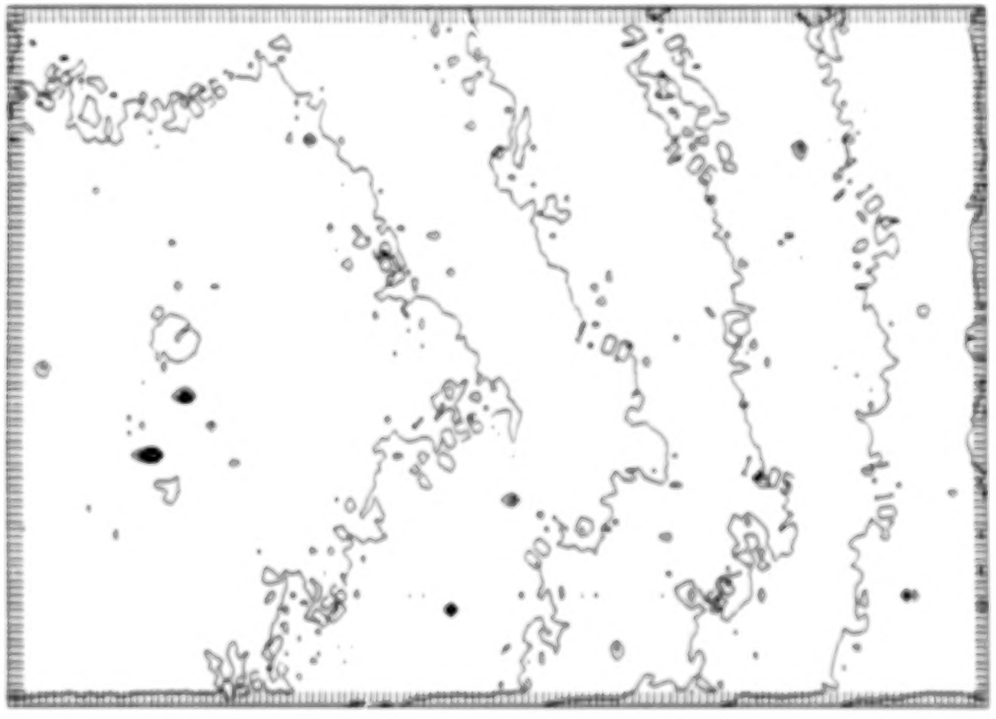


Figure 9. Relative illumination pattern of a projected lamp flatfield with respect to the sky median frame, for the Lick GEC CCD. The isophotes are labeled with illumination ratios.

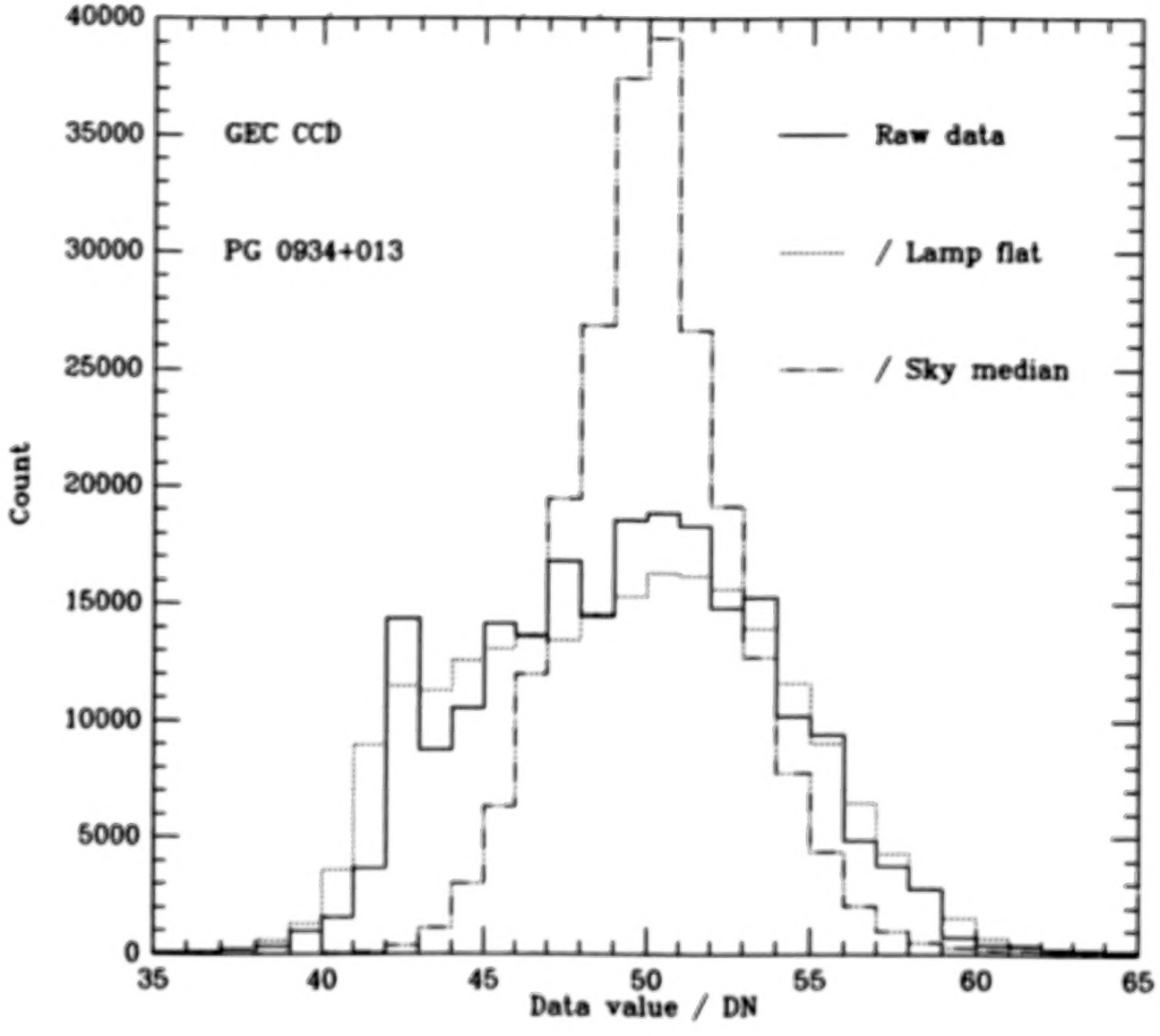


Figure 10. The efficiency of flatfielding can be judged through tightness of the data histograms before and after flatfielding. Solid line represents raw data for an exposure of QSO PG 0934+013 field with the Lick GEC CCD. Division by a simple incadescent lamp flatfield (dotted line) makes things slightly worse. Division by a sky median frame (dash-dotted line) improves on the scatter considerably.

the data with intrinisic color drastically different from the sky color, the flatfielding is not as effective.

If an adequate median-for-the-run is not available, an incandescent lamp flatfield can be used instead, but its performance is not as good, because of the color dependence of high-frequency sensitivity variations.

The efficiency of the flatfielding can be judged by the tightness of data histograms, as in Fig. 10. Inadequate flatfielding (e.g., straightforward use of incandescent lamp flats, or dawn-sky flats) can sometimes make things worse, and increase the total noise. Poissonian noise of the flatfield is always added to that of the data. The flatfielding is profitable only if the pattern noise of the chip is larger than the Poissonian noise of the flatfield, and the color differences between the data and the flatfield are small.

8. Low-End (Deferred Charge) Nonlinearity

Some of the charge deposited in a CCD is never read out; a minimal threshold amount is required for an effective charge transfer. The effect of this deferred charge is to lower the response curve of a CCD by a fixed amount in the high-signal regime, and make it highly nonlinear in the very-low-signal regime. This is schematically illustrated in Fig. 11. The effect is occasionally refered to as "fat zero" or "skinny zero" (which just serves to perpetuate the confusion). A detailed treatment of this problem is given in Baum, Thomsen & Kreidl (1981).

The amount of deferred charge, Δ, ranges from ~ 0 to several tens of electrons per pixel, for different devices. To a first approximation it can be treated as a constant for the chip, to a second approximation as a constant for a given column. A simple way to determine the intercept Δ is to obtain several frames of a same object (e.g., a lamp flatfield), with graded exposure times, and perform a least-squares solution on the frames. As a result, one obtains a frame of intercepts, whose modal data value is the estimator of Δ. An example is shown in Fig. 12.

The Δ for a given chip appears to be very stable in time, and does not depend perceptibly on the wavelength. In some columns, Δ can even have a small positive value (due to charge diffusion?).

Once the value of Δ is determined, the simplest way to cure this nonlinearity is to add the constant to every image before flatfielding:

$$\{\text{flattened image}\} = \{\text{raw image} + \Delta\} \,/\, \{\text{flatfield} + \Delta\}$$

A somewhat better way of correcting for this effect in the "toe" of the response curve is described by Baum, Thomsen & Kreidl (1981). Define a frame function μ:

$$\mu = \{\text{raw image}\}/\{\,|\text{raw image}| + c\}$$

where $|...|$ denotes absolute value and c is a constant to be determined. Then the flatfielding is accomplished as:

$$\{\text{flattened image}\} = \{\text{raw image} + \mu\cdot\Delta\} \,/\, \{\text{flatfield} + \Delta\}$$

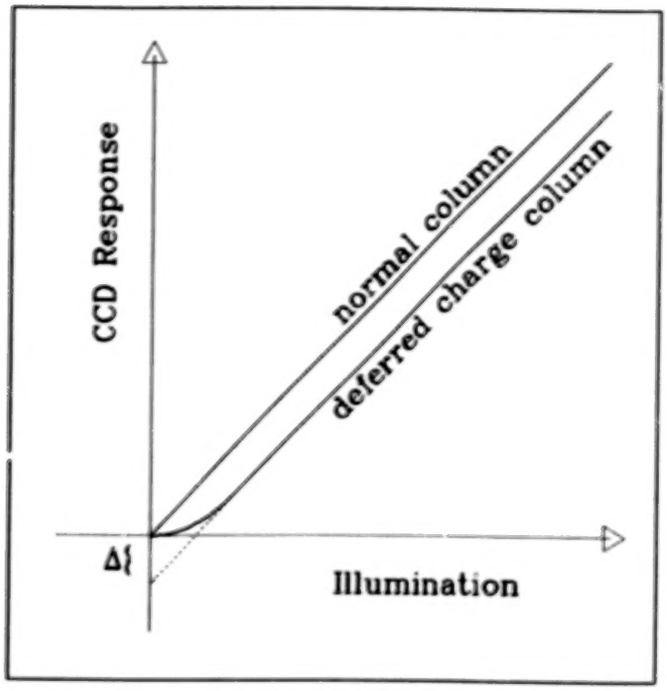


Figure 11. Schematic illustration of the deferred charge effect (after Baum, Thomsen, & Kreidl 1981).

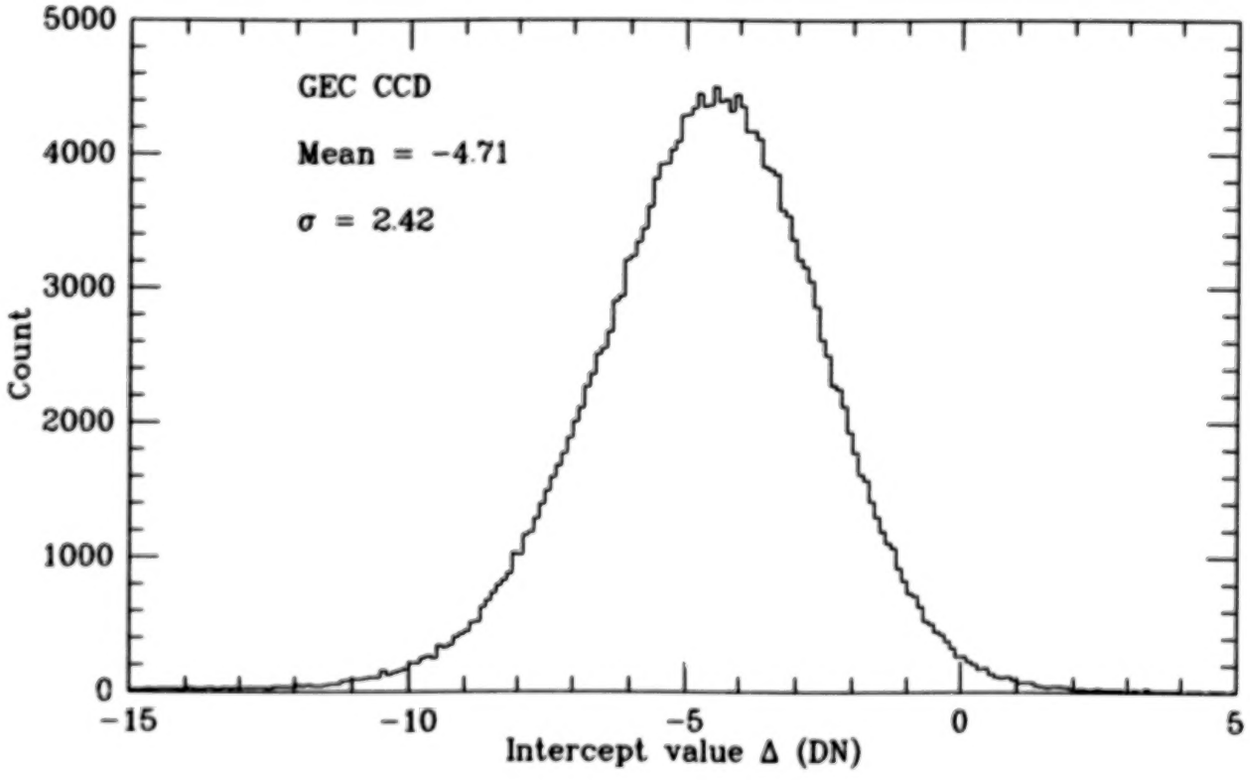


Figure 12. Data histogram for the frame of deferred charge intercepts for the Lick GEC CCD. Some of the scatter is due to poissonian noise. The frame was obtained from a sequence of graded-exposure flatfields, via least-squares method.

9. Cosmic Rays and Cosmetic Defects

CCDs are excellent cosmic-ray detectors. The cosmic-ray (CR) events appear as bright spikes in the data, and they are mostly associated with secondary muons.

In the old, thick (bulk) devices, CR events resembled stars or faint galaxies; they were of multi-pixel, low-level type. Multiple exposures were necessary in order to remove statistically such spurious "objects" (Goad 1980; Leach & Gursky 1979). In contrast to that, in the modern thin devices the CR events are predominantly of high-signal, single-pixel variety (Fig. 13), viz., non-stellar in appearance. Grazing-incidence events are seen, but they are rare, and also clearly non-stellar and easy to recognize.

Here we have a piece of good luck: the energy spectrum of CRs falls off as a power law on the high-energy side, whereas the dewar + telescope + dome shielding truncates it exponentially at the low-energy end. Thus, the charge spectra of detected CRs are peaked, in the well-recognisable regime. We do not have to suffer the nightmare of data pollution by many low level events. Detected charge spectra of CRs are shown, e.g., in Wright (1982), and a very instructive comparison of RCA thick and thin devices is shown in Fowler, Waddell & Mortara (1981).

Various cosmetic defects are seen in virtually all presently used CCDs. They include bad or blocked columns or parts thereof, bad pixels (pinholes), hot spots and hot pixels (which sometimes even emit light!), charge sinks, noisy edges, sensitivity drops, etc. They are mostly very obvious, and always present. Some defects are wavelength dependent, typically more prominent in the blue. Some are signal dependent. Since all of these phenomena are caused by physical defects or impurities on the CCD chip, their pixel locations are always the same. Thus, it is enough to map the chip defects once, and store their positions.

Both CRs and cosmetic defects can be treated in the same way. One either flags all inflicted pixels and does not use them when doing photometry, or interpolates across them by using the good data surrounding the flagged pixels. The first solution is more "honest", and it is easy to implement in the photometric schemes which use spatial information (e.g., PSF or isophote fitting), but not in others (e.g., pseudo-aperture photometry). The second scheme is generally prefered in purely imaging applications. Bad (noisy or warped) edge regions are best left unused in any application.

10. Fringing

The thickness of a CCD's active silicon layer is comparable with the wavelength of incoming light. If there are any deviations from a perfect planar geometry, or thickness variations over the chip, some interference pattern will be inevitably present. The night-sky emission features will produce the more prominent interference fringes. Those from the continuum will tend to average out. If the chip is mechanically fixed and stable (and they mostly are), the interference-fringe pattern will be spatially stable on the chip. However, since the night-sky emission lines and bands vary in relative intensity with time, the pattern is not stable in time. This can be particularly severe in the red/IR region.

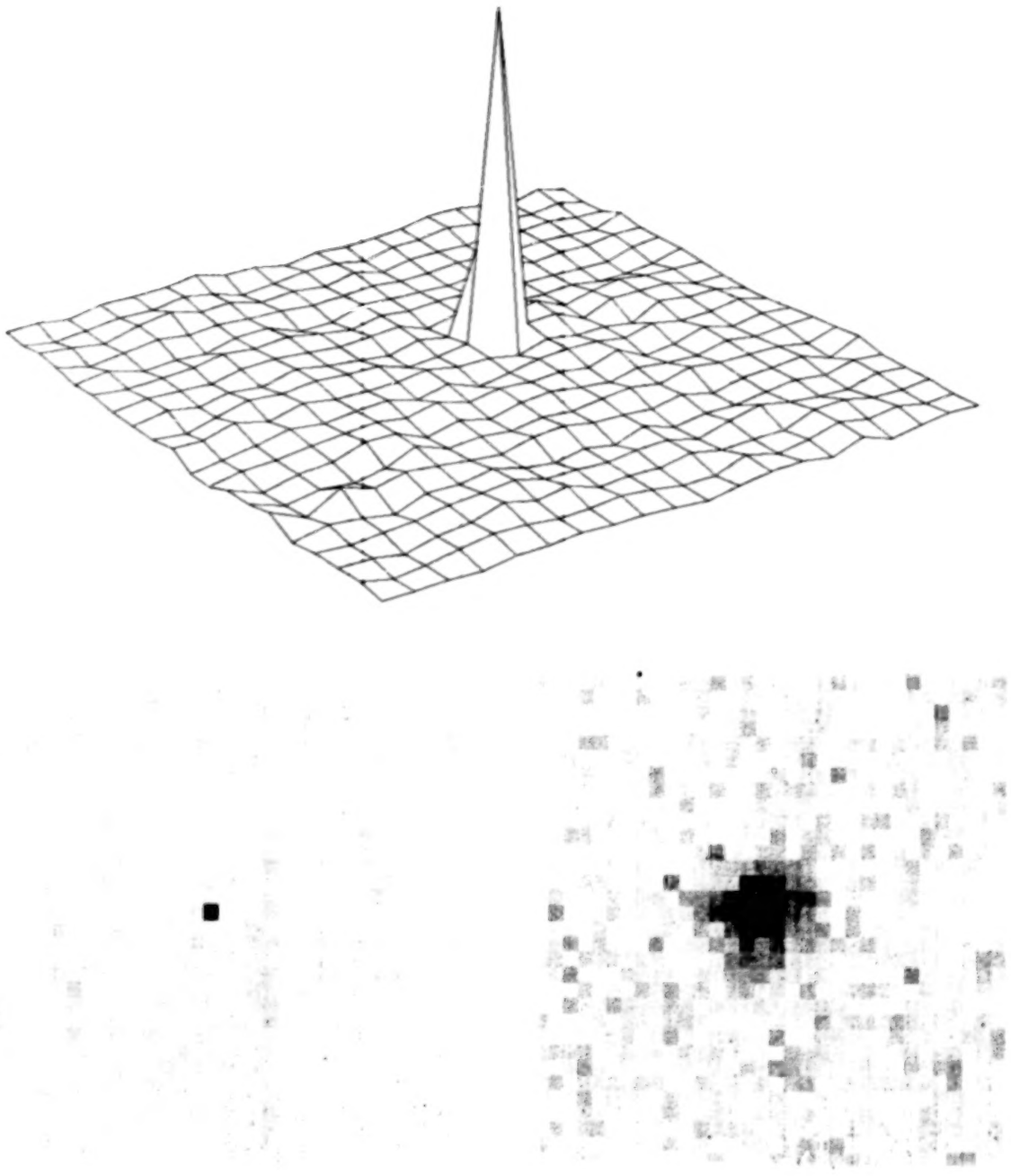


Figure 13. (Top) Surface plot of a CR event in the Kitt Peak Cryocam TI 800x800 CCD. This is a "1½" pixel event, standing well above the noise. (Bottom left) Gray scale plot of a single CR event in the Kitt Peak PF TI 800x800 CCD. (Bottom right) A faint star from the same CCD exposure, with about the same peak brightness, taken in good seeing. The morphological distinction is clear.

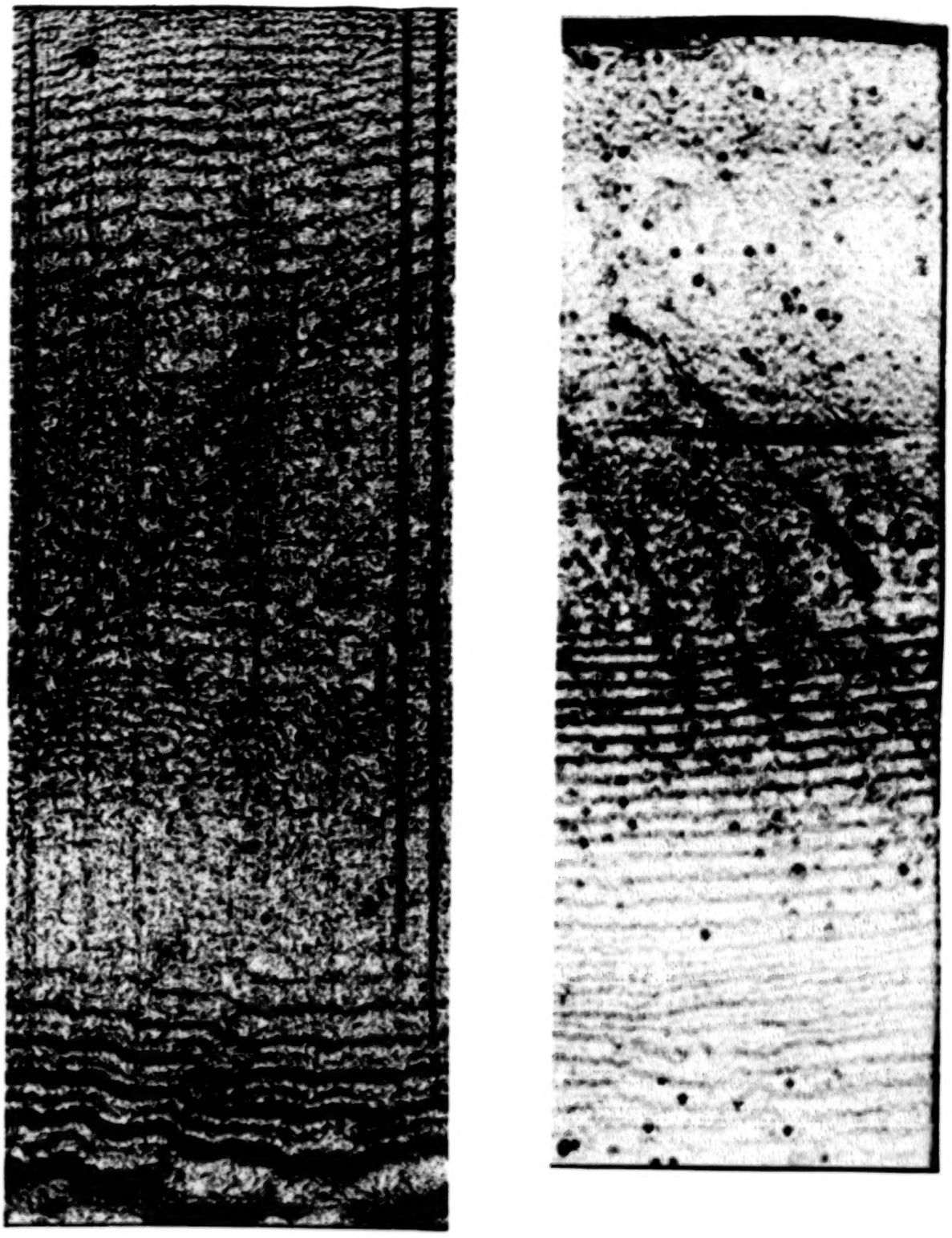


Figure 14. Spectroscopic CCD exposures (flatfields), enhanced with a two-dimensional gradient (Laplacian) filter. Interference fringes, hot spots, bad columns, and other defects are evident when this type of a spatial filter is applied. (Left) A 300x800 section of a Kitt Peak Cryocam TI 800x800 CCD infrared exposure. (Right) A 180x500 section of a Lick Cassegrain spectrograph TI 500x500 CCD visible-light exposure.

The fringe pattern is an additive signal (in imaging applications, but not in the spectroscopy), so it does not flatten out. For severly afflicted CCDs, it is useful to compile a library of object-pruned sky frames, from which one can make a linear combination, and subtract it from the data.

Fortunately, most modern devices are not severely afflicted with fringing, especially in broad-band applications. If one is interested in the photometry of of small objects (stars, faint galaxies) and the background is fitted and subtracted locally (as in PSF fitting photometry), fringing can be ignored. However, this luxury is not always available.

In spectroscopic applications, the fringes are wavelength-anchored, and if the physical pixel position is fixed with respect to the wavelength, the fringes flatten out. Examples of spectroscopic fringed frames are shown in Fig. 14.

11. Drift Scanning - a Panacea ?

The drift scanning technique is probably the best presently known way for exacting optimal CCD performance. The technique was first developed by Wright, and is described in Wright (1982), Wright & Mackay (1981) and Mackay (1982). It was also independently developed at Steward Observatory by McGraw, Angel & Sargent (1980).

The principle of drift scanning is simple: the CCD is moved with respect to the sky image, and it is read out at the same pace, so that an object image in the focal plane and its detection charge in the CCD move together until finally read out. The CCD motion is accomplished either by a moving stage, or by letting the telescope lag in tracking, or even stay fixed in azimuth. The CCD readout rate is adjusted so as to match the sky image motion across the surface. The output images are arbitrarly long strips.

In effect, every object is sampled by all the pixels in the appropriate readout columns. Many nonuniformities, nonlinearities, intra-pixel sensitivity variations, fringes, etc., are averaged along the readout columns. The only flatfielding remaining to be done is between the readout columns, and it is easily accomplished on the data directly. There is no need for a special flatfield frame, and all the pitfalls described in Sect. 7 are avoided.

This technique should come very close to the Poisson limit, and it is claimed (Wright 1982) that it indeed does. Effective flatfielding accuracy of the order of 10^{-4} is currently discussed. The second major benefit is the averaging over low-level systematics. If the data are fully Poisson-limited, then one can push the accuracy of the measurement by sufficient repetition of exposures. Note that these advantages appear only in the high-accuracy regime; for applications where the required accuracy is $> 1\%$ of the total signal (object+sky), the benefits of drift scannng are not as prominent.

Drift scanning is at the present still somewhat experimental. It appears very promising, but most stringent tests and actual "limiting" applications are still in the future.

12. Physical Limits to CCD Photometry

We can now summarize the limits in achieving high photometric accuracy with present technology CCDs. We discuss here stellar photometry only; surface photometry is more difficult and has other limiting factors. We also disregard here all external sources of error: transparency variations, dust, telescope imperfections, etc. The limits to relative accuracy quoted here are the CCD limits only.

For a good CCD in a non-scanning mode, in a single, sky-limited exposure:
- The Poisson limit is ~ 10^{-3} / star. This assumes a well-sampled image and a typical well capacity of ~ $5 \cdot 10^4$ electrons.
- The local sky determination can be good to ~ 10^{-3} (or maybe slightly better). This assumes negligible or well-removed fringing.
- Flatfielding RMS is typically ~ 10^{-2} / pixel, or ~ 10^{-3} / star.
- Systematic flatfielding errors are ~ 10^{-3}.
- Local deviations from linearity are ~ 10^{-3} (but they can be greater, if charge transfer problems are important).

Thus, the overall limit for a single exposure is about $(2\text{-}3) \cdot 10^{-3}$ per star, and for multiple exposures perhaps ~ 10^{-3} / star. This accuracy plateau is determined primarly by the low level systematic errors. In the limit of many exposures with different placements of the star on the chip, one approaches the results of drift scanning, and may do even better.

For a good CCD in a drift scanning mode, in a single sky-limited exposure:
- The Poisson limit is the same, ~ 10^{-3} / star.
- The local sky determination is Poisson-limited itself, and should be good to ~ 10^{-3} or better.
- The efective flatfielding should be good to ~ 10^{-4} or better.
- The systematics should mostly average out, but it is not known at the present time just how much.

The total limit is ~ 10^{-4} / star. With most systematics gone, multiple exposures should be able to push the accuracy, in $\sqrt{N}$ fashion for at least another decade down, and perhaps even more. Multiple exposures with objects sampled by different columns will help average the systematics even more.

CCDs have peak quantum efficiencies of the order of unity; no big improvements can happen there. Increased physical size and wider spectral response are clearly desirable. But when it comes to high-precision photometry, we should hope for more spatially uniform response, better linearity, and better charge transfer properties. This is also very important for the space-based observations, which will as a rule have very low background levels. When technology advances bring improvements for those aspects of CCDs, we can hope to push the level of photometric accuracy to the Poisson limit over an arbitrary magnitude range.

References :

Amelio, G. F. 1974, "Charge-coupled Devices", Sci. Am. February 1974, p. 23.

Baum, W. A., Thomsen, B., and Kreidl, T. J. 1981, "Subtleties in the Flat-fielding of Charge-coupled device (CCD) Images", Proc. SPIE 290, 24.

Blouke, M. M., Janesick, J. R., Hall, J. E., and Cowens, M. W. 1981, "Texas Instruments (TI) 800x800 Charge-coupled Device (CCD) Image Sensor", Proc. SPIE 290, 6.

Cawson, M. 1983, "Fast Shape Analysis of Galaxies", Occasional Reports of Royal Obs. Edinburgh #10, 175.

Djorgovski, S., and Spinrad, H. 1983, "Long-Slit CCD Spectroscopy of Very Faint Objects", Proc. of the Joint Topical Meeting on Information Processing in Astronomy and Optics, St. Paul, Minn., AAS/OSA special publication.

Fowler, A., Waddell, P., and Mortara, L. 1981, "Evaluation of the RCA 512x320 Charge-coupled Device for Astronomical Use", Proc. SPIE 290, 34.

Geary, J. C., and Kent, S. M. 1981, "Imaging Characteristics of the RCA 512x320 Charge-coupled Device (CCD)", Proc. SPIE 290, 51.

Goad, L. E. 1980, "Statistical Filtering of Cosmic Ray Events from Astronomical CCD Images", Proc. SPIE 264, 136.

Goad, L. E., and Ball, W. F. 1981, "Charge-coupled Device (CCD) Detector Utilization and Development at Kitt Peak National Observatory (KPNO)", Proc. SPIE 290, 130.

Gursky, H., Geary, J., Schild, R., Stephenson, T., and Weekes, T. 1980, "Astronomical Performance of the 512x320 RCA Charge-coupled Device (CCD)", Proc. SPIE 264, 14.

Janesick, J. R., Hynecek, J., and Blouke, M. M. 1981, "Virtual Phase Imager for Galileo", Proc. SPIE 290, 165.

Kent, S. M. 1983, "CCD Photometry of the Center of M31", Ap. J. 266, 562.

King, I. R. 1983, "Accuracy of Measurement of Star Images on a Pixel Array", Publ. Astron. Soc. Pacific 95, 163.

Kristian, J., and Blouke, M. 1982, "Charge-coupled Devices in Astronomy", Sci. Am. October 1983, p. 67.

Lauer, T. R. 1983, "High Resolution Surface Photometry of Elliptical Galaxies", Ph.D. thesis, Lick Observatory, Univ. of California at Santa Cruz.

Lauer, T. R., Miller, J. S., Osborne, C. S., Robinson, L. B., and Stover, R. J. 1983, "CCD Use at Lick Observatory", Proc. SPIE 445, 132.

Leach, R. W., and Gursky, H. 1979, "The Cosmic Ray Background in Charge Coupled Devices", Publ. Astron. Soc. Pacific 91, 855.

Leach, R. W., Schild, R. E., Gursky, H., Madejski, G. M., Schwartz, D. A., and Weekes, T. C. 1980, "Description, Calibration, and Performance of a Charge-coupled Device Camera", Publ. Astron. Soc. Pacific 92, 233.

Lorre, J. J., and Gillespie, A. R. 1980, "Artifacts in Digital Images", Proc. SPIE 264, 123.

Mackay, C. D. 1982, "Drift Scan Observations with a Charge-coupled Device (CCD)", Proc. SPIE 331, 146.

McGraw, J. T., Angel, J. R. P., and Sargent, T. A. 1980, "A Charge-coupled Device (CCD) Transit-telescope Survey for Galactic and Extragalactic Variability and Polarization", Proc. SPIE 264, 20.

McGuire, T. E. 1983, "The Kitt Peak CCD Camera System", Publ. Astron. Soc. Pacific 95, 919.

Mortara, L., and Fowler, A. 1981, "Evaluation of Charge-coupled Device (CCD) Performance for Astronomical Use", Proc. SPIE 290, 28.

Newell, B. 1983, "Crowded Field Stellar Photometry", Occasional Reports of Royal Obs. Edinburgh #10, 15.

Robinson, L. B. 1981, "The Lick Observatory Charge-coupled Device (CCD) and Controler", Proc. SPIE 290, 124.

Tody, D. 1980, "Stellar Photometry in Crowded Fields", Proc. SPIE 264, 171.

Tody, D. 1981, "RICHFLD Photometry Program Users Guide", Kitt Peak National Observatory, Tucson, Az.

Williams, T. B., and Schwarzschild, M. 1979, "A Photometric Determination of Twists in Three Early-Type Galaxies", Ap. J. 227, 56.

Wright, J. P., and Mackay, C. D. 1981, "The Cambridge Charge-coupled Device (CCD) System", Proc. SPIE 290, 160.

Wright, J. F. 1982, "The Application of Imaging Charge Coupled Devices in Astronomy" Ph.D. thesis, University of Cambridge, England.

Young, A. T. 1974, "Photomultipliers: Their Cause and Cure", Methods of Experimental Physics 12a, p.1, New York: Academic Press.

Zucchino, P., Long, D., Lowrance, J. L, Renda, G., Crawshaw, D. D., and Battson, D. F. 1981, "Evaluation of RCA Thinned Buried Channel Charge-coupled Devices (CCDs) for Scientific Applications", Proc. SPIE 290, 174.

TIME SERIES PHOTOMETRY WITH A CCD CAMERA

Alistair R. Walker
South African Astronomical Observatory

Time series observations of a field in the cluster NGC 4755 with a CCD camera are used to show that differential photometry between stars on the CCD frame is limited in accuracy mostly by photon statistics, over a 5 magnitude range. Scintillation noise appears to be almost entirely suppressed. The maximun accuracy possible is limited by the storage capacity of the CCD to about 0.001 mag.

I. INTRODUCTION

It has been amply demonstrated (eg Mould et al. 1983) that on a given telescope CCD's can produce accurate photometry of faint stars better than photoelectric or photographic techniques. The advantage over photoelectric photometry is due to the CCD being able to measure the sky background better because of its 2-D format while the advantage over photographic photometry results from the CCD's linearity, high quantum efficiency and ease of calibration of its pixel to pixel response. Additionally, Walker(1984) has shown that accurately timed short exposures of bright photoelectric standard stars with a CCD can produce results indistinguishable from those achieved using conventional photoelectric techniques on the same stars, residuals of a few 0.001 mag. being typical.

The purpose of this paper is to demonstrate that the CCD can be used to advantage for monitoring or searching for variable stars with periods of the order of minutes to hours in the situation where perhaps several candidate stars can be included on a single CCD frame while at least one other star on the frame can be used as a local standard.

II. INSTRUMENTATION

Observations were made using the University College London - SAAO CCD camera on the 1.0 meter telescope at Sutherland. This camera uses an RCA SID53612 thinned, back-illuminated CCD as detector and has been described in detail by Walker et al.(1984). It consists of a hardwired sequencer which controls the various functions of the camera by interpreting commands from a NOVA 3/12 computer which

also operates the shutter, controls the preflash and monitors and displays status. Data are read into the memory of a Sigma ARGS 7000 display, viewed on a TV monitor and can optionally be written to 9 track magnetic tape in FITS format. The shutter operating arm position is directly monitored and the exposure time absolute error is less than 10 ms, this degree of accuracy is essential if short exposures of bright standard stars are to be attempted. An assembler language program operates the exposure sequence, controls the movement of data between the peripherals and displays status information. A repeat mode allows for an exposure to be initiated immediately following the readout of a previous exposure, without operator intervention. For full sized frames (512x320 pixels) the readout time plus system overheads total 22 seconds. However it is possible to readout the CCD as though it were 256x160 "super-pixels", which takes only 9 seconds, and has the further advantages of reducing the readout noise and conserving magnetic tape. On our 1.0 meter f/16 telescope these super-pixels correspond to 0.8 arc sec square so unless the seeing is very good star images will not be under-sampled. This mode of operation was used for the observations described below.

III. OBSERVATIONS

A field approximately 2 arc minutes S of the centre of the open cluster NGC 4755 was chosen for this experiment, and 150 30 second exposures, each separated by 9 seconds, were made through a V filter. The sky was only marginally photometric, seeing was 3-4 arc sec and there was noticeable image wander and scintillation. An autoguider was used.

IV. DATA REDUCTION

The CCD frames were cleaned in the usual manner (Walker 1984) with subtraction of bias and preflash followed by division by a flat field frame obtained from a short exposure on the twilight sky.

Star 1 (Arp IV-27, Arp & van Sant 1958) was first measured by digital aperture photometry using an aperture of 20 arc sec, with the sky measurement being taken from an annulus of equal area surrounding the star aperture. These data were then corrected for extinction and can be considered to be equivalent to measurement by a simultaneous star-sky photometer.

Six stars were measured by fitting Gaussian profiles. For each frame a point spread function (PSF) was established by finding the best fit Gaussian to the sum of the three brightest, non-crowded stars. These profiles were then

fitted to six stars chosen not to have any close companions. Although we find that Lorentzians are much better fits to star images than are Gaussians, the latter were used here to minimize computer time, however this precluded measuring any stars crowded by companions. This case can be handled by programs which we have that fit Lorentzians to several star images simultaneously. The magnitudes resulting from the Gaussian fits were then normalised to that of Arp III-17 (V=9.7).

V. RESULTS

The top panel of figure 1 shows the digital aperture photometry measurements of Arp IV-27. There is evidence for a change in transparency near the end of the data, while the std. deviation is 0.008 mag per point. The expected error from photon statistics alone is 0.0022 mag. The lower panel of figure 1 shows the light curve for Arp IV-27 using Arp III-17 as a local standard. Apart from a dramatic decrease in the noise level, the std. deviation now being 0.0029 mag., the transparency change near the end of the data has been removed. Since using Arp III-17 as standard introduces an error of 0.0009 mag. due to photon statistics, the noise level achieved in figure 1 is therefore close to that expected by considering photon statistics alone, indicating that the scintillation noise must have been mostly removed.

TABLE 1. - NOISE STATISTICS

STAR	V	--PHOTON STATISTICS-- star alone (mag)	with std. (mag)	OBSERVED ERROR (mag)
III-17	9.7	0.0009	---	---
IV-11	11.4	0.0019	0.0021	0.0030
IV-27	11.8	0.0022	0.0024	0.0029
III-17	14.0	0.0065	0.0066	0.0083
anon	14.1	0.0068	0.0069	0.0099
anon	14.4	0.0077	0.0078	0.0111

Table 1 summarizes the results for the six stars measured. For the fainter stars the difference between the achieved noise level and that expected by considering photon statistics alone is greater than that for the brighter stars, attributable to readout noise from the CCD becoming significant. For these six relatively bright stars sky noise is negligible. Flat fielding errors, which are mostly due to

photon statistics, are estimated at 0.4 percent per pixel for the flat field used here. However flat fielding errors are rendered negligible due to the autoguider preventing the star images from moving relative to the CCD frame, hence differences in brightness between stars on a given frame are not affected by this error source.

VI. CONCLUSIONS

It has been demonstrated that the CCD can produce high quality photometry when comparing stars on the same frame, and over a 5 magnitude range the dominant noise source is photon statistics from the stars and sky. The major disadvantage is that a local standard must be found on the frame, which would tend to restrict measurements to regions of high star density such as star clusters. However in this case many stars can be measured simultaneously and, given suitable fitting techniques, crowded or overlapping images can be photometered. The ultimate limitation is the storage capacity of the CCD (about 240000 electrons per pixel) which makes it difficult to store more than 1-2 million electrons per star without saturating, the precise value depending on the pixel scale and the seeing. The maximun accuracy possible thus appears to be about 0.001 mag.

REFERENCES

Arp, H.C. & van Sant, C.T., 1958. Astron. J., 63, 341.

Mould, J.R., Kristian, J. & da Costa, G.S., 1983. Astrophys. J., 270, 471.

Walker, A.R., 1984. Mon. Not. R. astr. Soc., in press.

Walker, D.D., Walker, A.R. et al. In "Proceedings of the Eighth Symposium on Photoelectronic Image Devices", ed. B.L. Morgan, publ. Academic Press, London, in press.

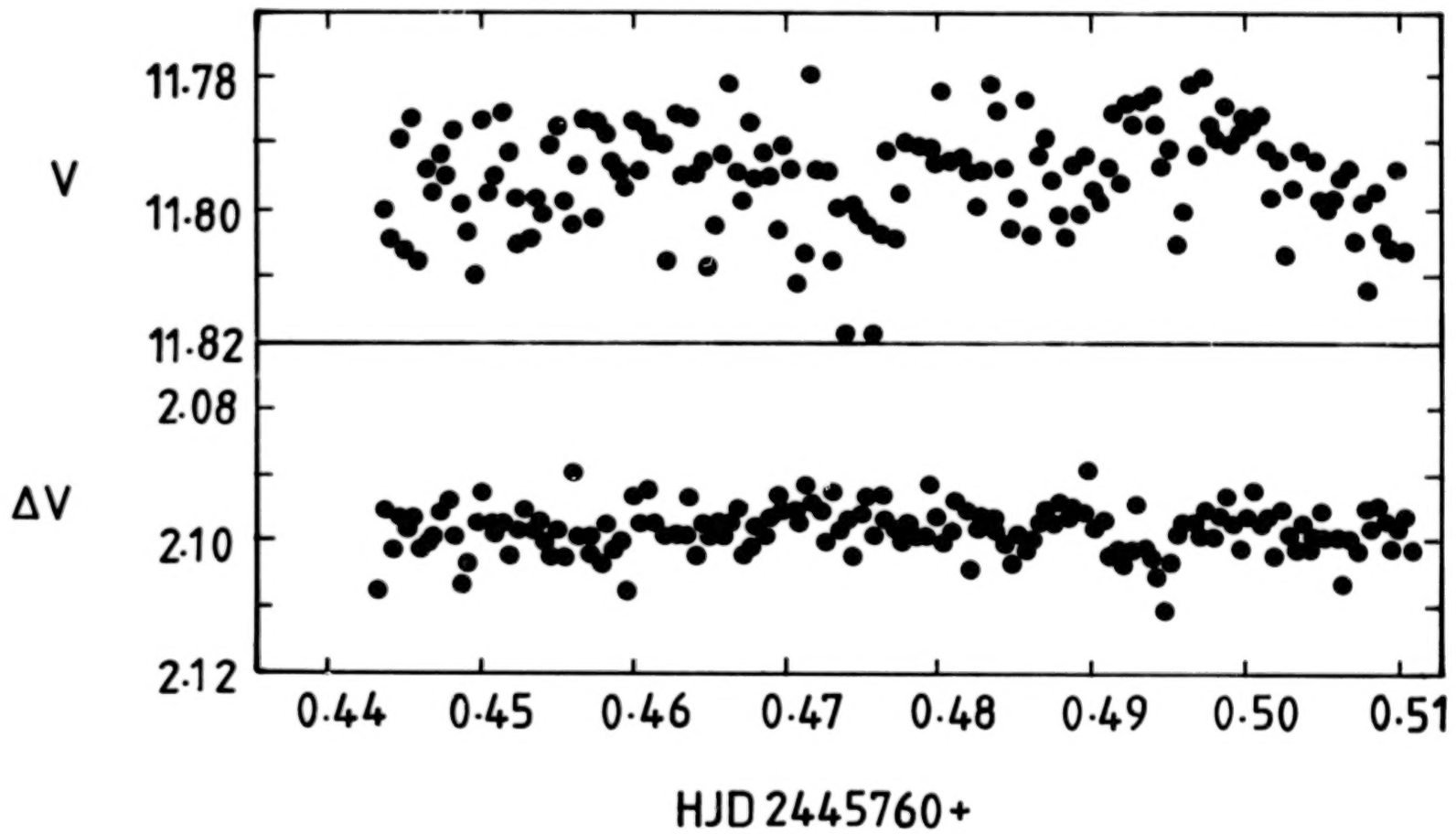


Figure 1.- Digital aperture photometry measurements of Arp IV-27. The lower panel of figure 1 shows the light curve for Arp IV-27 using Arp III-17 as a local standard.

N85
17907

UNCLAS

D15

The Use of Self-Scanned Silicon Photodiode Arrays for Astronomical Spectrophotometry

Anita L. Cochran
The University of Texas at Austin

Abstract: This paper discusses the use of self-scanned silicon photodiode arrays for precision spectrophotometry. It is shown that internal errors are ±0.003 mag. Observations obtained with a photodiode array are compared with observations obtained with other types of detectors with agreement, from 3500 A to 10500 A, of <1%. The photometric properties of self-scanned photodiode arrays are discussed. Potential pitfalls are given.

I. Introduction

Spectrophotometry is probably one of the most exacting types of astronomical observations. It requires excellent sky transparency, good seeing, and a stable instrument. In order to make spectrophotometric observations as accurately as possible, great attention must be paid to the details of the spectrograph and to the detector. This paper will discuss one type of detector, a Reticon Corporation self-scanned silicon photodiode array, in use at McDonald Observatory, and the kind of high quality results we have been able to obtain.

II. Spectrograph Design

In order to measure absolute light levels from an astronomical object, photometric techniques are used. Broad-band filters, such as UBV or uvby, can give an indication of color, metalicity and surface gravity. Higher spectral resolution is needed to obtain more detailed information about a star. In the past, this higher resolution has most often been achieved through multi-filter photometry or a grating scanner with a photomultiplier tube (PMT) as a

detector.

The multi-filter photometer is generally a single PMT and some number of discrete filters through which the star is imaged sequentially. This method has certain photometric advantages: linear response; high quantum efficiency compared to photographic plates; can use photon counting techniques; accurate sky subtraction; low noise; simple electronic systems; ease of data reduction. This method also has disadvantages: limited dynamic range (many photometers cannot be used to observe the primary standards); poorly defined bandpasses; bandpass stability of filters affected by temperature; often must use multiple PMTs to cover the wavelength range so observations cannot all be made in a single night; lack of multiplexing.

Spectrophotometry can be done with a spectral scanner. Typically, a star is imaged through a relatively large focal plane aperture and the light is dispersed by a grating. With a single PMT, the grating must be scanned to change wavelength. This technique is capable of giving much higher spectral resolution than the multi-filter photometer. However, this technique suffers from a lack of multiplexing resulting in an extreme loss of efficiency, and the data are sensitive to variations in signal due to atmospheric seeing and scintillation and to guiding errors. These problems can be minimized but not eliminated by rapid scanning. In addition, the wavelength accuracy of a scanner is mechanically limited.

The problems of mechanical scanning are overcome with modern multiplexed detectors. Gains in speed over a system with only one detector are obtained by a factor proportional to the number of detector elements in the system. The ideal multi-channel detector should have high quantum efficiency, linear response, no dead space between detective elements, broad wavelength coverage, and large dynamic range. The Reticon Corporation self-scanned silicon photodiode array (SPD) comes close to meeting this ideal.

The Reticon SPDs come in various linear models with 256 to 4096 diodes apiece. Some models (the "S" series) have height/width of 100:1. Some models (e.g. the CP1001 series chip) have two parallel arrays which can be used for accurate sky subtraction. The peak responsive quantum efficiency (RQE) is about 80% from 7000 A to 9000 A (Vogt, Tull, and Kelton, 1978 hereafter VTK), and is greater than that of any conventional photocathode over the entire wavelength region from 0.3 - 1.1 μm (VTK; Talmi and Simpson, 1980 hereafter TS). There is, effectively, no dead space between diodes and these diodes have extremely high geometric accuracy, leading to high wavelength precision. TS have shown this accuracy to be quite high. VTK(see their figure 12) have shown Reticon SPD arrays are linear over a range of at least 2.5×10^4, or 11 stellar magnitudes. TS tested the reciprocity of these arrays using a pulsed LED (see figure 11 of TS) and conclude that the reciprocity over a range of 10^6:1 was linear to <3% (the accuracy of their test equipment).

Thus, it would seem that Reticon SPDs are nearly ideal detectors for spectrophotometry. In section III, results of spectrophotometric observations

obtained at McDonald Observatory will be presented. In section IV, some of the problems which have been experienced with these detectors will be discussed.

III. Spectrophotometric Observations from McDonald Observatory

Reticon SPDs have been in use at McDonald Observatory for both spectroscopy and spectrophotometry since the early 1970's (Tull and Nather, 1973; Stover, 1976; VTK). The first spectrophotometer was a small, low resolution instrument with a Reticon Corporation RL256C/17 SPD (Stover, 1976). It contained 256 diodes on 25.4 μm centers. Each diode has an active area of 420 x 25.4 μm. The grating provided a reciprocal dispersion of 820 A/mm resulting in 42 A resolution (2 diodes) and wavelength coverage from 4650 A to 10200 A.

The instrument had two entrance apertures which were alternately imaged onto the detector by wobbling the grating perpendicular to the dispersion. Thus, half of the time was spent observing the star and half observing sky. This permitted accurate sky and fixed pattern signal (see section IV) removal. Though half of the time was spent observing sky, this was not an onerous burden since observations time were short. On the 0.9 m telescope, the spectrum of an M-type star with V=10 mag could be obtained in about 90 seconds (45 seconds on star and 45 seconds on sky) and the spectrum of 10th mag A-type star could be obtained in about 10 minutes (stars fainter than 10th mag could not be viewed in reflection from the slit with the 0.9 m telescope). Figure 1 shows typical uncalibrated spectra from this instrument.

This instrument was used from 1977 August 30 to 1978 April 21 to obtain spectrophotometric observations of stars. The instrument was treated as a 256 channel photometer for purposes of data reduction and standard techniques were used to compute extinction coefficients and instrumental response at each diode.

Cochran and Barnes (1981) investigated the photometric accuracy of this instrument by obtaining repeated, independent observations of α Lyr on several nights. They found an unexplained systematic shift of the zero-point of their observations which gave a standard deviation in the zero-point of a single observation of ±0.025 mag. The mean of the 15 observations deviate from the standard α Lyr spectrum by ±0.006 mag (standard error of the mean) (fig. 2).

To determine the internal precision, Cochran and Barnes normalized the 15 observations near 6000 A and recomputed the standard deviation. The standard deviation of a single observation was then ±0.003 mag leading to an expected

standard deviation in the mean of the 15 observations of about ±0.0008 mag.

Cochran (1981) used this instrument to recalibrate the secondary spectrophotometric standard stars directly against the bright primary standard, α Lyr. Many instruments cannot be used to observe α Lyr. Thus, many stars in the past have been calibrated using a boot-strap technique. Our Reticon observations were then compared with previous spectrophotometric observations of these stars. Cochran found that, with the exception of the region of the Paschen jump (see paper for details) her observations agreed with those of Hayes (1970) and Taylor (1979) to 0.008 ± 0.004 mag. This is well within the errors caused by the zero-point shift.

Failure of the temperature control circuitry for this spectrometer in April 1978 resulted in the spectrometer being removed from use. In April 1981, a 2x936 CP1001 Reticon SPD was installed on a conventional Cassegrain spectrograph for use on the McDonald Observatory 2.1 m telescope. This detector has two parallel arrays of 936 diodes each. The active area of each diode is 30 x 375 μm with no dead space between adjacent diodes or between the two arrays. The spectrograph has two entrance slits mapping onto the two arrays. There is a quartz wobble block above the slit. The star is first imaged through slit 1 onto array 1 while the sky is imaged onto array 2. Then the wobble block is tilted to image the star through slit 2 onto array 2 while array 1 views the sky. Thus, we alternately obtain two independent sky-subtracted spectra of the star.

These observations are reduced using standard spectrophotometric techniques and the data from the two arrays are kept separate during the reduction for comparison at the end. Thus, we treat this detector as a system of 1872 separate photometers.

Wavelength coverage for this spectrograph is user selectable by changing gratings and grating angles. Wavelength regions from 0.3 μm to 1.1 μm may be observed at resolutions (2 diodes) from 1.5 to 14 A. With these set-ups, one can still observe the primary spectrophotometric standard, α Lyr, even on a 2.1 m telescope. We also routinely observe objects as faint as 12.5 mag. At a resolution (2 diodes) of 6.5 A, we routinely achieve signal/noise of 100 per diode at V for an exposure of 30 minutes, for each of the two independent sky-subtracted spectra.

Using this instrument, Neff et al. (1984) observed Uranus, Neptune and Titan in the spectral range 0.35 to 1.05 μm at 7 A resolution. Exposures ranged from 2 minutes to 10 minutes (per slit), depending on which object and which spectral range. They converted the observations to fluxes and geometric albedos, and compared them with the scanner data of Lockwood et al. (1983) for Uranus. Agreement between the Neff et al. and the Lockwood et al. observations was better than 1% at all wavelengths.

In addition to the Reticon SPDs at the Cassegrain foci, we have long used

Reticon arrays at coudé. W. Cochran (personal communication) obtained spectra of Jupiter and Venus using a Reticon SPD with 1728 diodes at the coudé spectrograph of the 2.1 m telescope in April 1980. He was searching for the 0-6 band of HCl which has a bandhead at ~6321 A. On Jupiter, he obtained four spectra with signal/noise (SNR) ranging from 500 to 850 for an overall SNR of ~1450. The resolution element was 0.060 A. None of the expected lines of HCl were visible. (A 3σ feature would have a depth of 3/1453 and a width of 0.060 A.) This non-detection leads to an upper limit for the equivalent width of one of these lines of 0.12 mA.

For Venus, six observations were obtained with SNR of 1000 each. The SNR of the sum is then 2550. Again, no line was detected. This leads to an upper limit of the equivalent width of 0.07 mA. Clearly visible interstellar lines with equivalent widths of as small as 0.15 mA have been positively detected with a similar system (Snell et al.,1977).

IV. Potential Pitfalls

We have shown that the Reticon SPDs are capable of high precision spectrophotometry, but they are not without potential pitfalls. These pitfalls have been discussed extensively in the literature (cf. Livingston et al., 1976 hereafter LHST; VTK; TS) but it is worth discussing a few of them here.

Reticon SPDs are not photon-counting devices. They are composed of discrete silicon p-n junction photodiodes. Each is initially charged by a reverse bias of 5 volts. Photons arriving at the diode decrease its charge at a rate of one electron-hole pair (e-h pair) per absorbed photon. The recharge pulse necessary to restore the diode to its original 5 volts is proportional to the time integral of photon flux since the last readout, and becomes the signal measurement. In the absence of error due, e.g., to pre-amplifier noise, then this would be a perfect system. The most important sources of noise are preamplifier readout noise and thermal leakage shot noise. These sources of error can be dealt with and will be discussed individually below. Other potential sources of error in the spectrograph system will be discussed.

IV.1 Fixed Pattern Signal

In the absence of any other signal, each diode produces a characteristic zero offset; the fixed pattern signal of the array is the overall pattern of

these individual zero offsets. This is caused by capacitive coupling of the transients of the complementary clock driver signals onto the video lines. The level of this fixed pattern signal can be minimized by careful adjustment of the relative phases and amplitudes of the square clock waveforms but it cannot be fully eliminated. It is typically 1 - 5% of the full scale signal and it is repeatable from observation to observation if temperature is controlled and can, thus, be calibrated out by a simple subtraction. By taking a separate dark or sky observation, the fixed pattern signal can be nearly completely removed; the residual is primarily due to readout noise.

IV.2 Readout Noise

A complete discussion of readout noise is given by VTK. Some readout noise is generated from pickup on the clock, start, bias, and video lines; in a well engineered system, these sources are dominated by noise of the front-end transistor of the pre-amplifier. The readout noise sets the limiting brightness in spectrophotometry. The overall noise equivalent charge for Reticon SPDs has been found to be 800 - 1000 e-h pairs (rms) per readout (LHST; VTK; TS), equivalent to an uncertainty of 800 - 1000 detected photons in each diode.

For high signal levels, photon shot noise tends to dominate; for low signal levels, it is necessary to reduce the effective readout noise prior to reading out. The total readout noise is proportional to the square root of the number of readouts whereas the total photon flux increases proportionally with integration time. Thus, single long integrations have become routine for faint objects. Thermal leakage tends to become more important with long integrations.

IV.3 Thermal Leakage

In the absence of any form of signal, thermally induced currents slowly discharge the diodes. The rate of discharge changes by approximately a factor of 2 for each 6.7°C temperature change (TS). At room temperature, the array will be completely discharged in about 10 seconds (VTK). VTK (see their figure 14) have found thermal leakage rates of <1 electron/sec at temperatures below -125°C. Therefore, cooling the array can reduce the thermal leakage to a very small source of error. At temperatures below -125°C, where the thermal leakage is almost negligible, thermostating to ±1°C is adequate. At warmer temperatures, temperature control must be more rigorous to not allow changes in the rate of thermal leakage. Holding temperature constant has the added advantage of maintaining the fixed pattern signal as a constant.

At McDonald Observatory, the cooling system used with the Cassegrain spectrograph on the 2.1 m telescope is a fairly simple one: the array is enclosed in a vacuum chamber built by Frank Melsheimer (DFM Engineering, Boulder, Colorado) with the array in thermal contact with a cold finger immersed in liquid nitrogen. This system will cool the array to --130°C, somewhat dependent on the ambient temperature. To control the temperature setting, a resistor used as a heating element is attached to the cold finger near the array. A temperature sensing diode monitors the temperature, which is controlled to about ±1°C. As a check on the stability of the temperature control, several diodes on each end of the array are masked to produce a dark reference level for all observations.

A cooled array experiences a large drop in RQE in the near ir. Figure 6 of VTK illustrates this point. At 1 μm, the RQE of a diode at -60°C is only 82% that of a diode at -30°C; at -115°C it is 60%; and at -130°C it has fallen to only 57% of the RQE at -30°C. The array is mostly unaffected by cooling shortward of 7000 A. Thus, there is a trade-off between cooling the array to reduce the thermal leakage and RQE in the red. A system where the temperature can be actively, rather than passively, controlled is, therefore, an advantage.

IV.4 Other Reticon Pitfalls

There are two important effects brought about by the properties of silicon. The transparency of silicon increases in the near ir, increasing past 1 μm. As the silicon becomes transparent, a growing fraction of the light can completely penetrate the silicon and reflect from the gold substrate below. This will cause interference fringe patterns to be produced, loss of signal and loss of resolution. LHST and VTK have noted these patterns in the near ir. They found that this effect was more severe in some chips than others and was non-existent in some chips. Careful selection of chips could eliminate the problem. However, even if the problem exists in a particular array, LHST found that if the detector was illuminated by monochromatic light, such as with a spectrograph, this fringe pattern could be totally calibrated out. The technique of determining the instrumental response diode-by-diode automatically handles this problem.

The degradation of resolution towards the near ir results from the increased transparency to long wavelength photons. IR photons penetrate deeper into the substrate before giving up their energy to e-h pair creation. This allows some amount of lateral diffusion of the e-h pairs before reaching the diode depletion level.

IV.5 Spectrograph Pitfalls

There are two important spectrograph-related effects that should be considered in all spectrophotometry. In order to pass all of the starlight from the telescope to the detector, spectrophotometers are typically slitless. Apertures must be available which are much larger than the seeing profile. This can cause two problems.

In a slitless grating spectrophotometer, the seeing disk will actually set the instrumental resolution. With changes in seeing, the resolution changes. In continuum light, this is unimportant as long as the continuum has only a small slope. Within a spectral line, the situation is different, resulting in changes in the shape of the line profile. This is, of course, a primary reason for using high dispersion spectrographs rather than low dispersion spectrometers in line profile studies.

In a slitless system, mis-centering or mis-guiding of the star image in a direction parallel to the dispersion will cause the resultant spectrum to be shifted. This can be compensated only imperfectly in data reduction. The only complete cure for this problem is accurate centering and perfect guiding. Since most telescopes track better in declination than right ascension, the problem can be minimized by mounting the spectrograph on the telescope so that the E/W axis is perpendicular to the dispersion.

V. Summary

Since spectrophotometry requires such exacting technique, the design of the instrument must be considered very carefully. Of paramount importance is the type of detector utilized. We have shown that an ordinary spectrophotometer equipped with a Reticon SPD is very nearly the ideal detector for accurate spectrophotometry of bright objects which can produce high signal/noise observations. There is a definite point at which observations with a Reticon SPD become useless -- that is, when the thermal leakage or readout noise begins to dominate the signal.

We would like to stress that, even with the best spectrometer and detector, the limit to the accuracy which can be obtained is in the observing technique used and, ultimately, in the quality of the standard stars. As more astronomers work at higher spectrophotometric resolutions, they will find that stars calibrated with broad band filters are not well enough calibrated for higher resolution work. With excellent detectors such as the Reticon SPDs, it

is imperative that we now work on obtaining the best calibrated observations possible.

References

Cochran, A. L. 1981, Ap. J. Suppl., 45, 83.

Cochran, A. L. and Barnes, T. G., III 1981, Ap. J. Suppl., 45, 73.

Hayes, D. S. 1970, Ap. J., 159, 165.

Livingston, W. C., Harvey, J., Slaughter, C. and Trumbo, D. 1976, Appl. Opt., 15, 40.

Lockwood, G. W., Lutz, B. L., Thompson, D. T. and Warnock, A. 1983, Ap. J., 266.

Neff, J. S., Humm, D. C., Bergstralh, J. T., Cochran, A. L., Cochran, W. D., Barker, E. S., and Tull, R. G. 1984, Icarus, in press.

Snell, R., Tull, R. G., vanden Bout, P., and Vogt, S. S. 1977, in Proceedings Special IAU Session, Grenoble, France, 30 August 1976, ed. J. Audouze(Dordrecht, Holland: D. Reidel) , 85.

Stover, R. J. 1976, Masters thesis, The University of Texas at Austin.

Talmi, Y. and Simpson, R. W. 1980, Appl. Opt., 19, 1401.

Taylor, B. J. 1979, A. J., 84, 96.

Tull, R. G. and Nather, R. E. 1973, 1973 Symposium of Astronomical Observations with Television-Type Sensors, Vancouver, B. C..

Vogt, S. S., Tull, R. G., and Kelton, P. 1978, Appl. Opt., 17, 574.

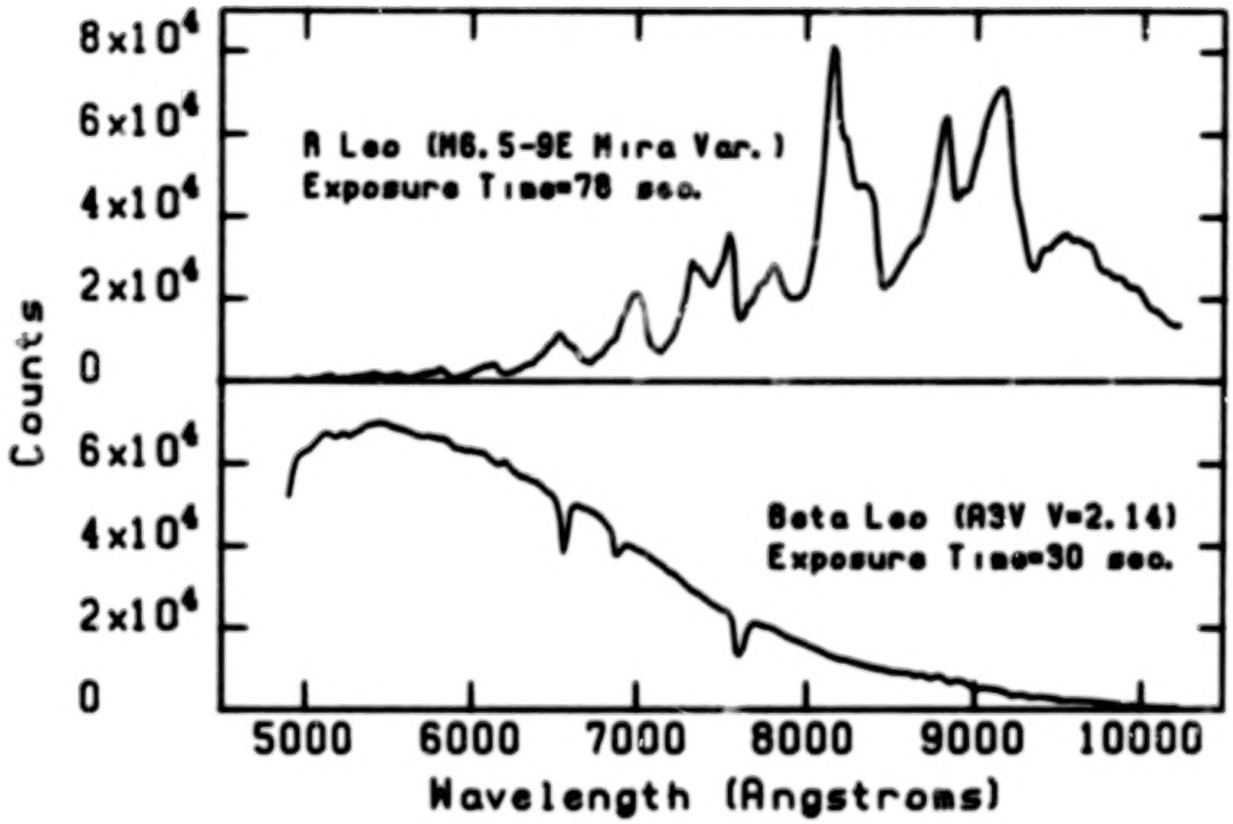


Figure 1.- Sample spectra taken with the 0.9 m Reticon spectrometer.

These spectra are not calibrated.

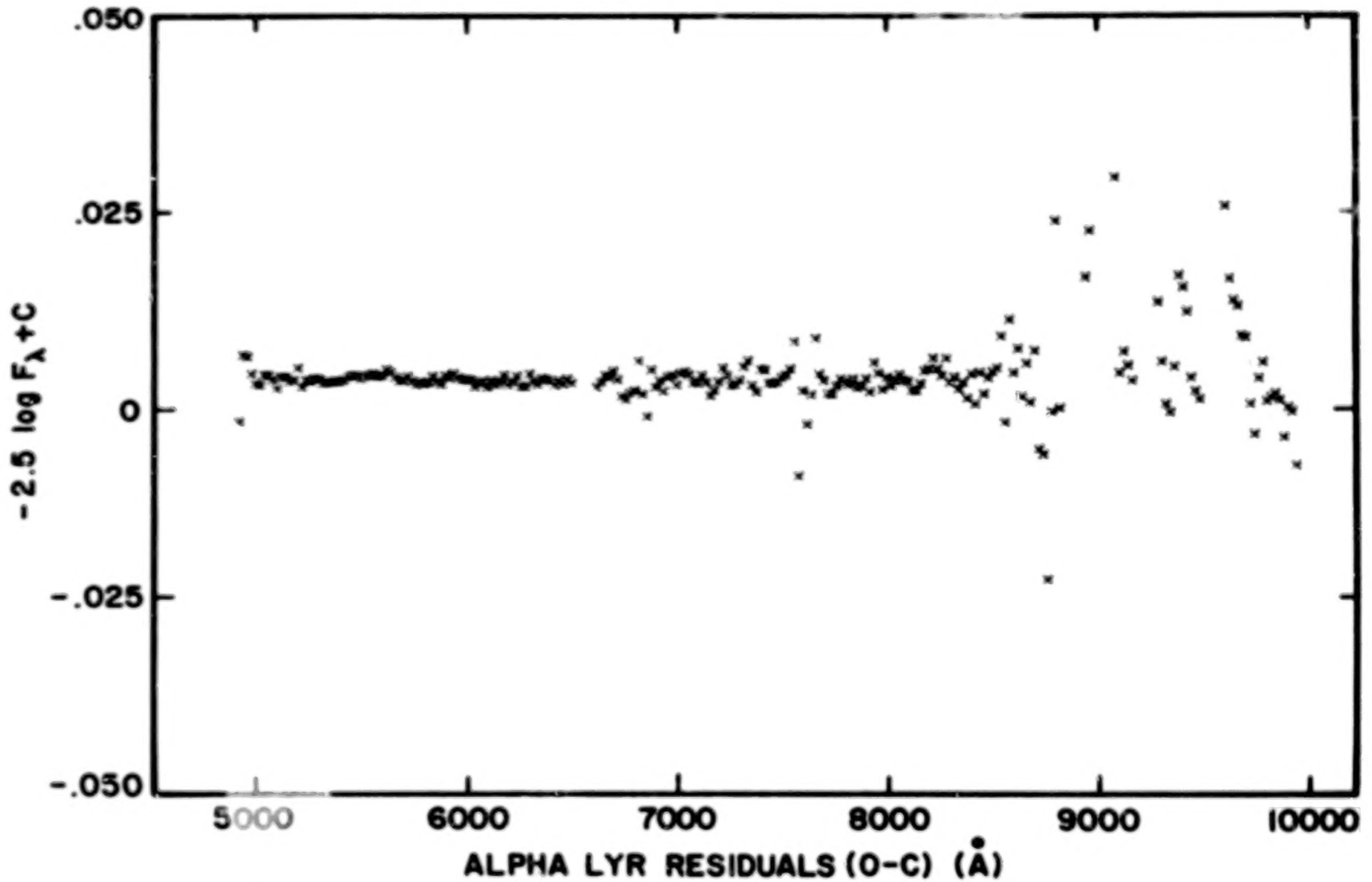


Figure 2.- Diode-by-diode residuals (in the sense observed minus standard calibration) for the mean of the 15 α Lyr observations. The scatter in the near IR is due to the presence of the Paschen lines. These lines are not well calibrated. The offset from zero is less than the standard error of the mean and is, thus, not significant. (This figure taken from Cochran and Barnes (1981), with permission.)

N85
17908

UNCLAS

PHOTODIODES FOR ASTRONOMICAL STELLAR RADIOMETRY

A. R. Schaefer
National Bureau of Standards

In this paper we will consider the applicability of solid state photodiodes to stellar radiometric measurements. We will present a brief review of the theory and application of silicon photodiodes to radiometric measurements, followed by some suggestions concerning photodiode detectors of possible use in these measurements.

The original work describing the use of silicon photodiodes as absolute standards based on fundamental physical principals was published beginning in 1979. The key ideas are presented by Zalewski and Geist (1980), Geist, Zalewski, and Schaefer (1980), and Geist, Liang, and Schaefer (1981). A thorough discussion of the underlying physics of photon flux measurements with silicon photodiodes is given by Geist, Gladden, and Zalewski (1982). Finally, a description of a silicon diode device with 100% external quantum efficiency is presented by Zalewski and Duda (1983). One should note that this has been a very rapidly developing field of endeavor. Consequently, some of the references and details given above are now obsolete; however, most of the ideas, especially those presented in Geist, Gladden, and Zalewski (1982) remain valid.

Figure 1, taken from Geist et al. (1982), illustrates a normal situation when using a p^+ on n planar diffused junction silicon photodiode to measure incident radiation. As is discussed in that work, one can justify two approximations: that the quantum yield, η, for electron-hole carrier production does not depend on photon penetration depth, x, into the silicon; and that there is no absorption in the front oxide layer. In that case, one can represent the external quantum efficiency of the photodiode as

$$Q(\lambda) = \left(1-\rho(\lambda)\right)\cdot\eta(\lambda)\cdot F(\lambda) \qquad (1)$$

where $1-\rho$ is the photodiode absorptance, η is the quantum yield for carrier creation, that is, the average number of electron-hole pairs created per absorbed photon, and F is the collection efficiency factor, which describes the fraction of minority carriers that can be swept across the junction and create a photocurrent flow in an external short circuit before recombination can take place. The solid lines in Figure 2, ibid., show typical values of these three factors as a function of wavelength, λ. The quantum efficiency is simply related to the spectral response, $R(\lambda)$, by

$$R(\lambda) = Q(\lambda) \cdot \lambda / \kappa \qquad (2)$$

where κ = 1239.85 nm W/A. In order, therefore, to determine the spectral response of a detector at some given wavelength and angle of incidence (F and ρ are functions of angle of incidence), one needs only to determine the three quantities, F, η, and ρ.

The determination of these three quantities is a straightforward task for a sufficiently high quality photodiode. For the optical quality, silicon dioxide antireflection coating used in these diodes, ρ is entirely specular, and therefore quite easy to measure experimentally. This is a very thin (~100 nm) and pure coating, so there is negligible absorption loss in this layer. (Two millimeter thicknesses of pure borosilicate glasses with comparable boron concentration show no absorption in the visible and near-uv spectral regions.) In the visible region of the spectrum, the quantum yield, η, is exactly unity. In the near uv, interband Auger transitions from energetic minority carriers leading to impact ionization can increase this above unity (Geist and Wang, 1983); and in the far red, for heavily doped silicon, free carrier absorption can decrease the quantum yield to less than unity, but these effects are negligible for the case we are considering. (A detailed discussion of these considerations is presented in Geist et al., 1982).

The third factor to be considered is the behavior of $F(\lambda)$, the collection efficiency factor. First, one must keep in mind the behavior of the optical absorption coefficient of silicon as a function of wavelength. It is much greater in the blue spectral region than in the red. Therefore, higher energy photons (blue) are absorbed near the surface region of the diode, while those of lower energy (red) will penetrate deep into the bulk of the detector before being absorbed. Qualitatively, the value of F is reduced below unity in the blue spectral region by recombination of minority carriers at the silicon-silicon dioxide interface. This happens because of their attraction to this interface by a positive charge trapped in the oxide surface coating, resulting in a surface majority carrier depletion region. For most of the mid visible, the photons are absorbed in the depletion region near the metallurgical junction of the diode (the high field region produced by the dopant charge profile). In this region the minority carriers are swept by the field across the junction at a velocity such that their transit time is much less than their recombination lifetime. Finally, in the farther red spectral region, the minority carrier is created beyond the depletion region, so that it must diffuse toward the junction some distance before being collected. Some recombination in the bulk region will again reduce the value of F below unity. This behavior is indicated in Figure 2. Both of these reductions in collection efficiency (and therefore quantum efficiency) can easily be measured experimentally, as outlined by

Zalewski and Geist (1980). One needs only to apply a voltage to the surface oxide; forming a charge to counteract the effect of the charge trapped in the oxide. One can then observe the increase in photocurrent corresponding to the increased internal quantum efficiency. This can be compared to the quantum efficiency with no charge producing voltage present. In a similar manner one applies a reverse bias to the diode to extend the depletion region and counteract recombination loss in the bulk of the diode. Again one can observe the correspondingly higher photocurrent. These new levels of F are represented by the dotted line in Figure 2. By increasing each of these voltages until the photocurrent is no longer observed to increase, we can determine the voltages necessary to insure that F, and therefore the internal quantum efficiency, is unity. These measured changes in the internal quantum efficiency can then be applied as correction factors when the diode is used without electrical bias in radiometric measurements. In some cases, particularly at higher flux levels in the red spectral region, it is preferable to do radiometry with reverse bias applied during the measurements.

Strictly speaking, the quantum efficiency in the far blue spectral region does not quite reach unity after the initial leveling out of photocurrent with applied oxide charge, but continues to slowly increase with increasing oxide charge accumulation. Geist, Liang, and Vertebout are developing a semi-quantitative theory to explain this phenomenon, which amounts to a reduction of quantum efficiency of about one-half percent in the near ultraviolet, but is negligible for the mid visible spectral regions.

There have been a number of independent radiometric verifications of this predictable quantum efficiency (PQE) technique (often referred to as silicon self-calibration). Zalewski and Geist (1980) originally compared radiant power measurements of a He-Ne laser beam using both the PQE and an electrical substitution radiometer described by Geist et al. (1973). Their results agreed to 0.06%, well within the 0.3% accuracy of the radiometer. (The accuracy of the PQE measurement was believed to be within ±0.04%). Probably the most stringent intercomparison was done in 1981 by Zalewski and Key (private communication) at the National Physical Laboratory in England, in which both PQE and a high accuracy cryogenic radiometer capable of accuracy of 0.01% were used to measure the radiant power in a series of Kr ion laser lines. The overall agreement was within 0.05%. Conditions of the experiment limited the uncertainty to ±0.1%. Similar successful intercomparisons have been reported by others, such as Hughes (1982), who compared PQE measurements performed with both lasers and lamp sources to those obtained with a pyroelectric absolute radiometer, and Li Tong-Bao (1984), who has performed PQE with both p^+ on n diffused junction diodes

and also inversion layer diodes such as those to be described below. They were used to measure He-Ne beam powers, with results comparing to within 0.1% of those obtained using electrically calibrated radiometers.

More recently, an inversion layer photodiode of the type developed by Hansen (1978) was studied and found to have negligible surface recombination losses by Geist et al. (1981). This diode consists of moderately doped p-type material in which a shallow n inversion is induced by positive charge trapped in the silicon dioxide surface coating. The minority carriers (holes) generated near the surface are thus repelled toward the junction and away from the interface. This results in unity quantum efficiency in the blue spectral region and makes unnecessary the experimental application of an oxide bias. In using this diode, however, one must apply reverse bias in order not only to extend the depletion region and improve the red quantum efficiency, but to prevent the photocurrent from forward biasing the detector to the extent of interrupting its operation. This type of diode, when properly fabricated, is nearer to ideal for use as a radiometric standard. Zalewski and Duda (1983) have described a device utilizing four inversion layer photodiodes with outputs summed together, arranged in a light trapping configuration to produce a detector with 100% external quantum efficiency for light in one polarization in the visible spectrum. This device is shown in Figure 3 (from Zalewski and Duda, 1983). Since the reflectance of each detector surface at 45° for TM polarized radiation varies from 0.08 to 0.33 over 400 to 800nm, the seven surfaces encountered by a beam in this device result in almost total absorption (99.9%). This device has been very successful. However it has two shortcomings: the acceptance angle is very narrow, making it suitable only for highly collimated sources such as lasers, and the 45° reflectance for TE polarization is too high, e.g., almost 0.6 at 406nm which would yield a quantum efficiency of only 0.986. A new device is presently being developed, consisting of three larger area inversion diodes arranged in a 90° spiral configuration. These diodes are of a very high quality and have a closely optimized antireflection SiO_2 layer. As discussed by Zalewski and Duda (1983), this device should be insensitive to the polarization effects just described, and should also have a much larger acceptance angle, so that it could be used with extended sources such as standard lamps.

Fisher (1968) has already discussed the possible advantages of using solid state detectors in astronomical radiometric measurements. He examined two silicon diodes and compared their performance in the visible and near ir with S-1 photomultiplier tubes, noting the much higher quantum efficiency of the diodes. He also noted another important feature: the extremely linear response of photodiodes over seven orders of magnitude. More

recent experimental studies (Schaefer et al., 1983, Budde, 1979, 1982, and 1983) have verified the linearity of silicon photodiodes over eight decades within 0.1%, corresponding to about 20 stellar magnitudes. Fisher also noted the insensitivity of photodiodes to magnetic fields, and their high degree of stability. Zalewski (private communication) has seen a slight change in the blue spectral response ~0.5% over a time period of one year in some p^+ on n devices, but in the mid visible and red spectral regions this change is only on the order of a few hundredths of a percent. Changes in the ultraviolet in excess of one percent per year have been reported by some investigators (Fox and Key, Williamson, private communications). Such changes are not expected with inversion layer diodes, according to our current understanding of their performance. Nonetheless, in any situation where an accurate knowledge of the absolute response (quantum efficiency) is required, the PQE technique should be used.

Obviously, one of the most critical properties limiting use of a detector for astronomical measurements is its noise current. The theoretical treatment of noise is not a simple subject, and will not be dealt with here. For diodes of the type we have discussed, operated without reverse bias, most of the noise is simply thermal (Johnson) noise, which can be represented as

$$i_n^2 = [4kT/R_{sh}] \cdot df \qquad (3)$$

where k is Boltzmann's constant, T is the absolute temperature, R_{sh} in general is the current derivative of the voltage vs. current characteristic of the detector: in this case its shunt impedance, and df is the frequency bandwidth interval over which the noise current is measured. This comes from Nyquist (1928). Typical values for this noise current quoted by detector manufacturers for the large (1 cm^2) area silicon photodiodes with which we have worked are ~ $4 \cdot 10^{-14}$ amps/$Hz^{1/2}$ at 10 Hz. If we define noise equivalent power (NEP) as noise current divided by responsivity, this corresponds to a NEP of about 10^{-13} watts/$Hz^{1/2}$ throughout the visible spectrum. For smaller diodes with an active area ~ 1 mm^2, this figure is reduced by one order of magnitude. These figures are consistent with what we have observed in the laboratory, although we have not yet done a systematic study of noise performance of diodes. It is interesting to note that Fisher (1968) examined diodes with NEP values comparable to these, and calculated that, with an equivalent filter bandwidth of 5 nm, at the Cassegrain focus of a 224 cm telescope with an optical transmission of 0.8, at 550 nm the limiting visual magnitude at which silicon diodes could be used is ~14. He also illustrates the NEP advantage of two to three decades of silicon diodes over several S-1 photomultipliers.

Another possible candidate for low level radiometry is gallium arsenide phosphide. In contrast to silicon, this is a direct bandgap material with a bandgap of 1.8 eV, resulting in a sharp long wave cutoff of about 700nm, instead of tailing off at about 1100nm as is the case for silicon. However, these properties result in a higher shunt impedance and lower value of thermal noise. According to manufacturers specifications, this should result in a NEP about one order of magnitude smaller than for a silicon diode of the same given area. It should be possible to improve NEP for either type of detector by cooling techniques. It should also be noted that Budde (1983) has measured the linearity of some GaAsP diodes and found them to be useful over an even wider range (ten decades, or 25 stellar magnitudes) than silicon, primarily because of their improved NEP over silicon.

In conclusion, we have described a general method using silicon photodiodes for absolute radiometric measurements based on fundamental physics. We have also considered their applicability for low level astronomical applications, and the possibility of using GaAsP diodes for such measurements.

REFERENCES

Budde, W. 1979, **Appl. Opt., 18,** 1555.

Budde, W. 1982, **Appl. Opt., 21,** 3699.

Budde, W. 1983, **Appl. Opt., 22,** 1780.

Fisher, Richard 1968, **Appl. Opt., 7,** 1079.

Geist, J., Gladden, W. K., and Zalewski, E. F. 1982, **J. Opt. Soc. Am., 72,** 1068.

Geist, J., Liang, E., and Schaefer, A. R. 1981, **J. Appl. Phys., 52,** 4879.

Geist, J., Schmidt, L. B., and Case, W. E. 1973, **Appl. Opt., 12,** 2773.

Geist, J., and Wang, C. S. 1983, **Phys. Rev. B, 27,** 4841.

Geist, J., Zalewski, E. F., and Schaefer, A. R. 1980, **Appl. Opt., 20,** 3795.

Hansen, T. E. 1978, **Phys. Scr., 18,** 471.

Hughes, C. G. 1982, **Appl. Opt., 21,** 2129.

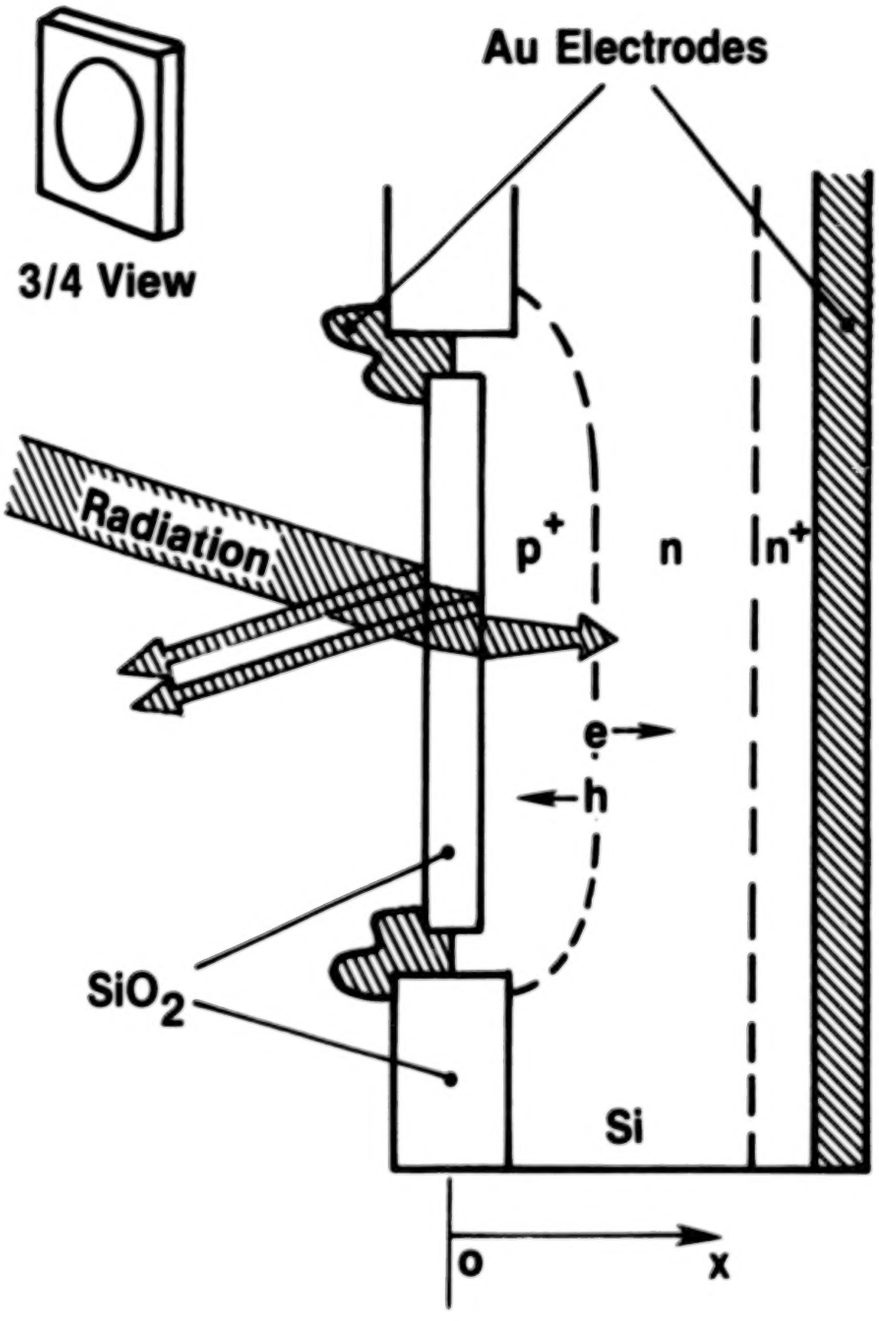
Li, Tong-Bao, and Sun, H. 1984, **Appl. Opt., 23,** to be published, July.

Nyquist, H. 1928, **Phys. Rev., 32,** 110.

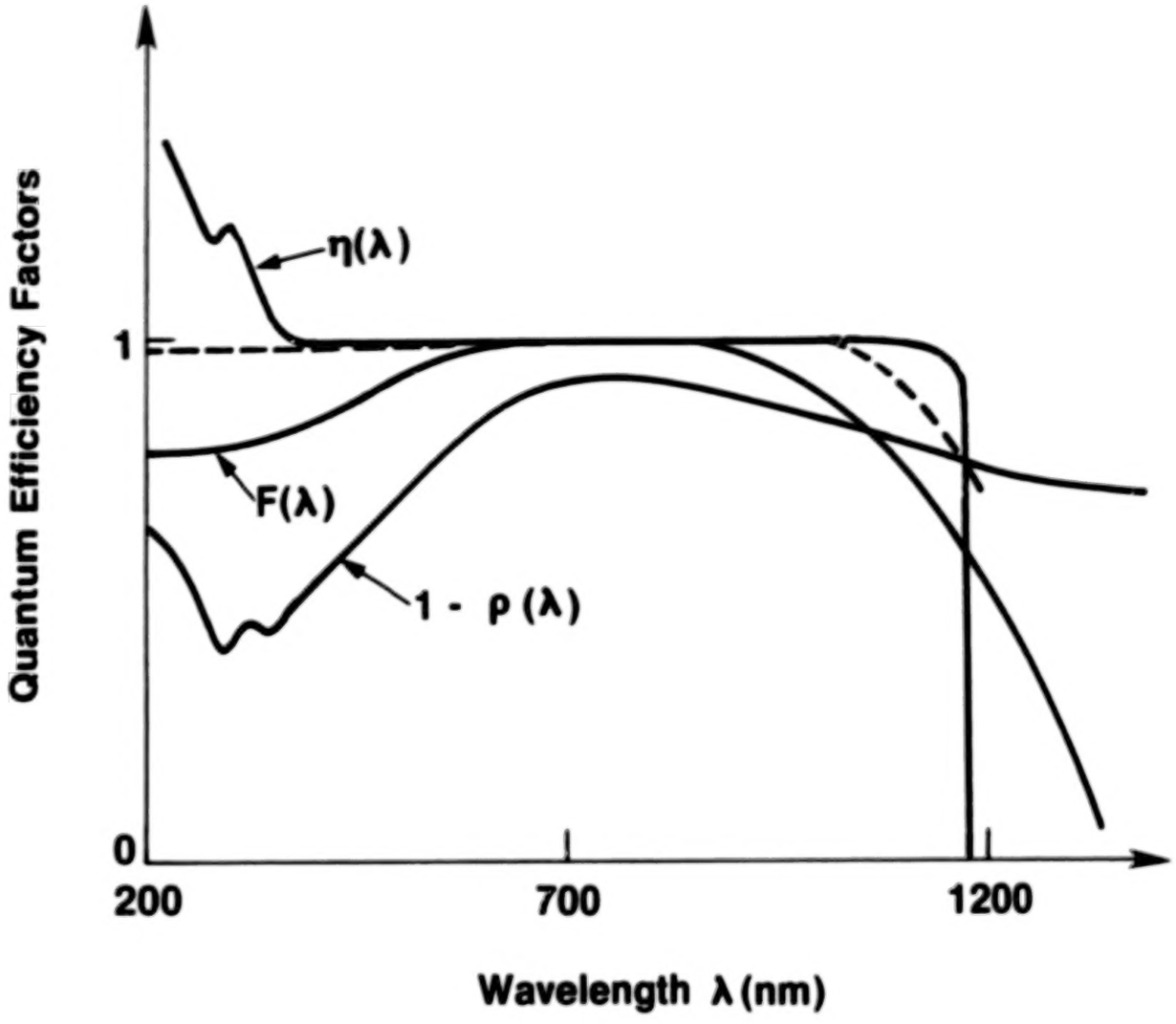
Schaefer, A. R., Zalewski, E. F., and Geist, J. 1983, **Appl. Opt., 22,** 1232.

Zalewski, E. F., and Duda, C. R. 1983, **Appl. Opt., 22,** 2867.

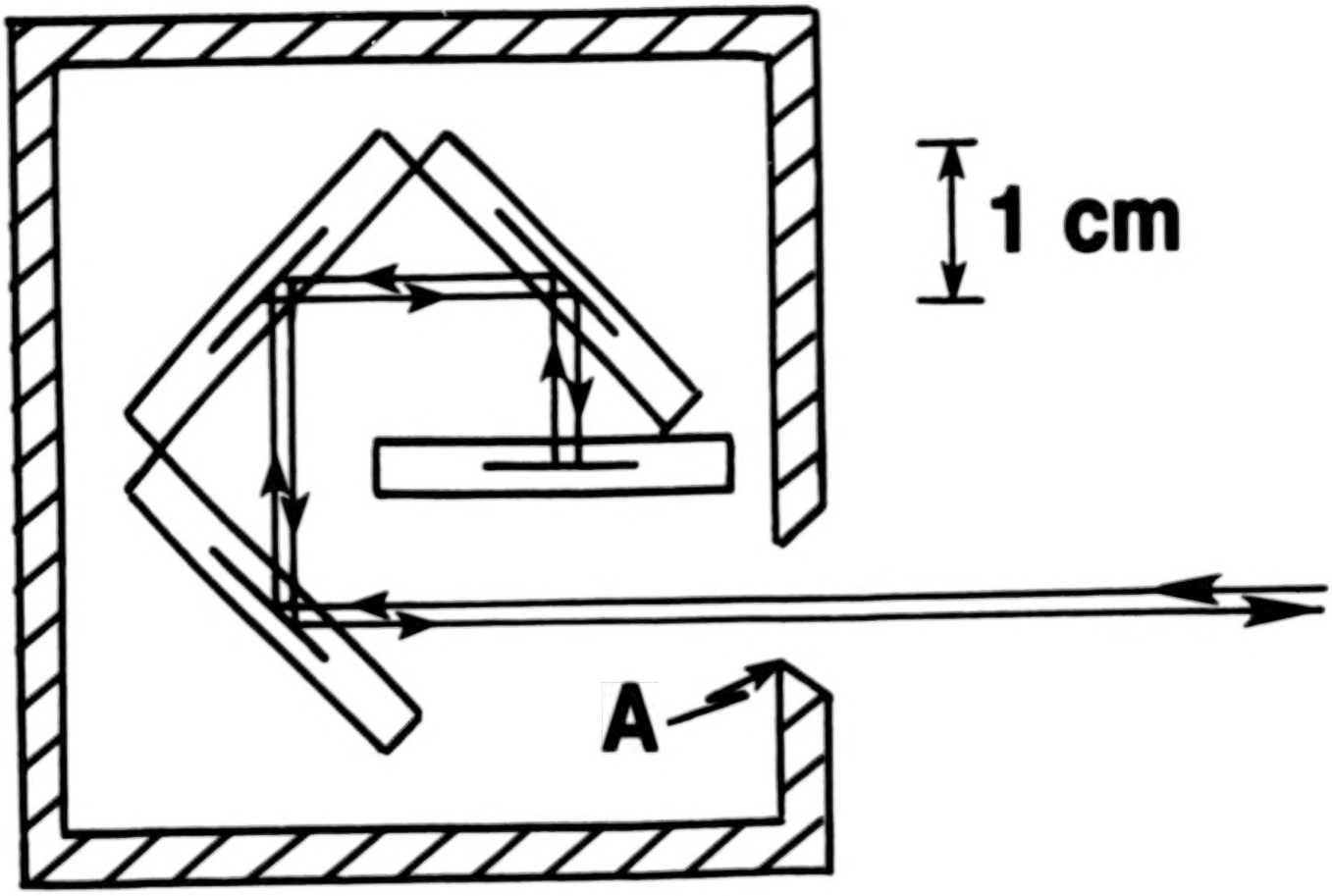
Zalewski, E. F., and Geist, J. 1980, **Appl. Opt., 19,** 1214.



1. A cross section of a p^+ on n planar diffused junction silicon photodiode, showing reflection and absorption of the incident radiation, with subsequent creation of electron-hole pairs.



2. Typical wavelength dependence of the three factors absorptance, $1-\rho(\lambda)$, quantum yield, $\eta(\lambda)$, and collection efficiency, $F(\lambda)$, whose product yields the external quantum efficiency, $Q(\lambda)$.



3. Optical configuration of 100% external quantum efficient detector. Reflected beam is shown displaced from incident beam for clarity.

Photometry with Multi-Anode Microchannel Arrays (MAMAs) and Charge Injection Devices (CIDs)

by

J. Gethyn Timothy

Center for Space Science and Astrophysics, ERL 315
Stanford University
Stanford, California 94305

The purpose of this article is to summarize the characteristics of two kinds of detectors with emphasis on those aspects that would affect their use in high-accuracy astronomical photometry.

(1) MULTI-ANODE MICROCHANNEL ARRAYS (MAMAs)

The Multi-Anode Microchannel Arrays (MAMAs) are a family of pulse-counting array detectors being developed by the autho at Stanford University in collaboration with Ball Aerospace Systems Division (BASD) specifically for use in space. The components of a MAMA detector system consist of a tube assembly (sealed or open) containing an anode array and a single curved-channel Microchannel Plate (MCP) with the appropriate photocathode deposited on, or mounted in proximity focus with, the front face (see Fig. 1), together with the associated amplifiers, discriminators, and address-decoding and counting circuits.

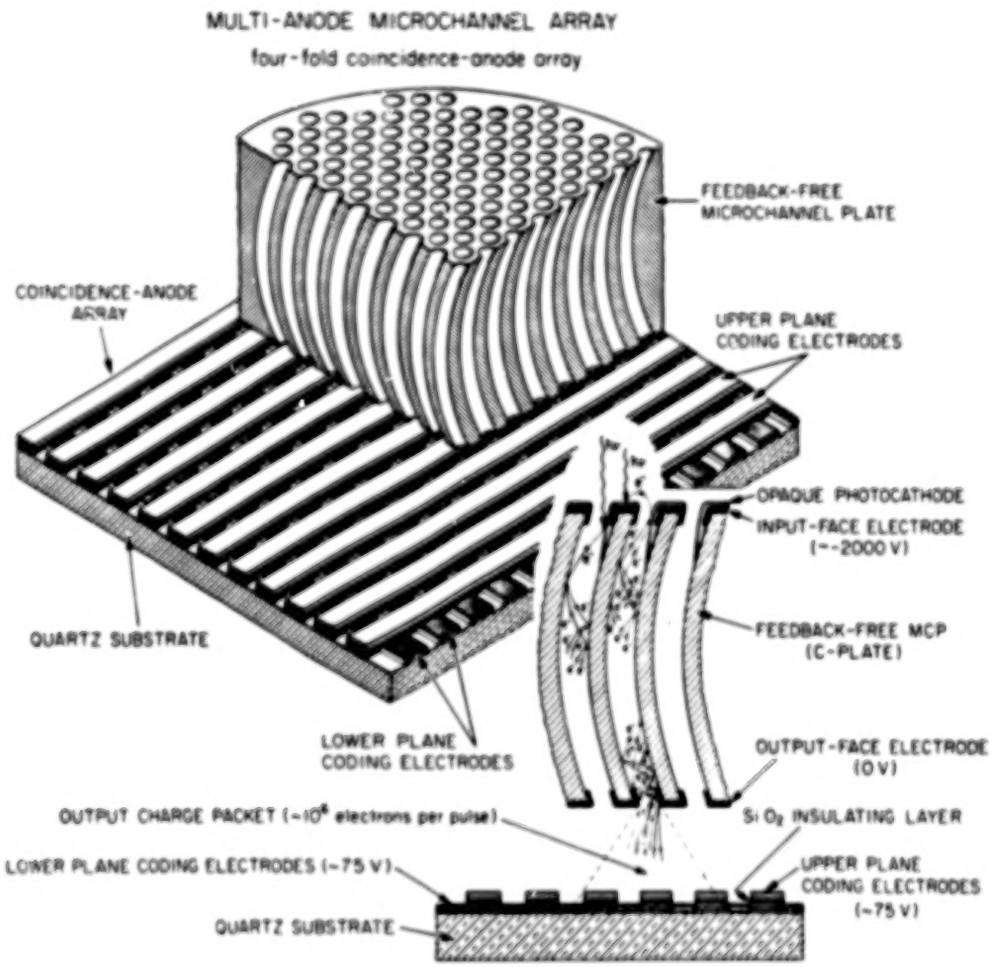
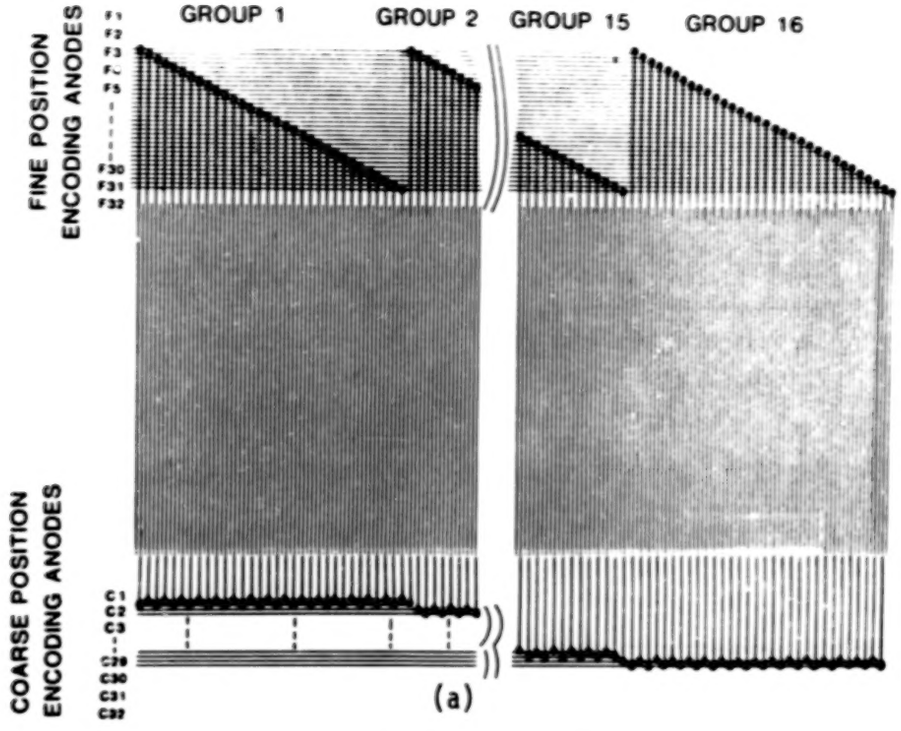


Figure 1. Schematic of Multi-Anode Microchannel Array (MAMA) detector.

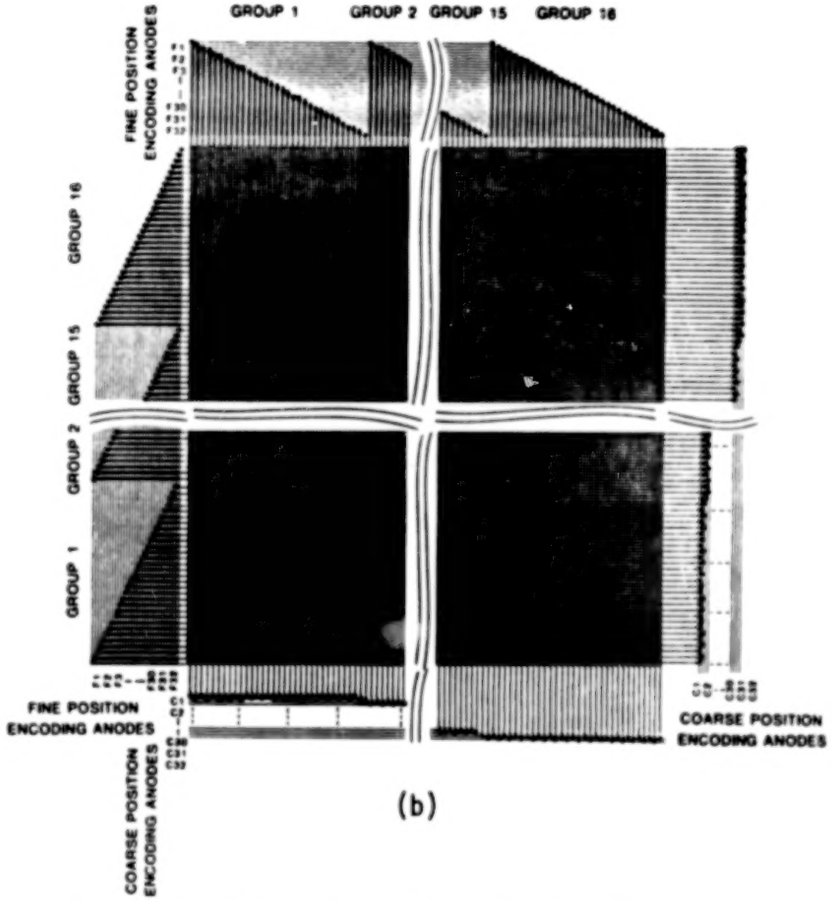
Two types of multi-anode arrays are currently under development. In the discrete-anode array (Timothy and Bybee 1981), each anode (i.e., pixel) is connected to an individual amplifier and counting circuit. In this type of readout system, the spatial resolution is defined solely by the geometry of the anode electrodes rather than by the charge-amplitude-division or timing-discrimination circuits that are employed in alternative MCP detector systems. The discrete-anode system however has a limitation in that the total number of pixels is restricted to about 500 by currently available connector and electronics technologies.

In order to overcome this limitation, the multi-layer coincidence anode array which employs two sets of anodes insulated from each other but exposed to the output face of the MCP, as shown in Fig. 1, is currently under development. The output charge cloud from the MCP divides between the two sets of anodes at the position where the event occurs allowing the spatial location to be identified by the coincident arrival of pulses on the appropriate pair of anodes. Photometric data from (a x b)-pixels can then be read out with a total of (a + b) amplifier and discriminator circuits. As shown in Fig. 2a, a (1 x 1024)-pixel linear array, for example, has a total of (32 x 32)-pixels and requires (32 + 32) amplifier and discriminator circuits.

Two-dimensional arrays can be constructed from two identical coincidence anode arrays arranged vertically above each other and oriented orthogonally. As shown in Fig. 2b, a (1024 x 1024)-pixel array fabricated in this manner employs coincidence detection in the two axes and requires only a total of 128 amplifier and discriminator circuits. The technical problem in the fabrication of these multi-layer arrays is the need to expose the lower-layer electrodes to the low energy ($\sim$ 30 V) electrons from the MCP which cannot penetrate an insulating layer. This is accomplished by etching away the silicon dioxide



(a)



(b)

Figure 2. Formats of one- and two-dimensional coincidence-anode arrays.
a) (1 x 1024)-pixel array.
b) (1024 x 1024)-pixel array (from Timothy and Bybee 1981).

insulating layer between the two sets of electrodes in the interstices between the upper-layer electrodes (Timothy and Bybee 1983).

Visible-light versions of the MAMA detector tubes employ a proximity-focused, semi-transparent photocathode as shown in Fig. 3. The 50 Å-thick aluminum oxide or silicon dioxide membrane on the front face of the MCP prevents photocathode degradation caused by back bombardment of positive ions formed within the MCP channels. The MAMAs are pulse-counting arrays, and unlike the CCDs or CIDs, do not store charge within the detector itself to be read out serially at some specified time interval. Rather, the MCPs are pulse-counting, random-readout detector systems. The arrival time and address coordinates of each detected photon event are strobed out in real time for storage on magnetic tape or for integration in a digital Random-Access Memory (RAM).

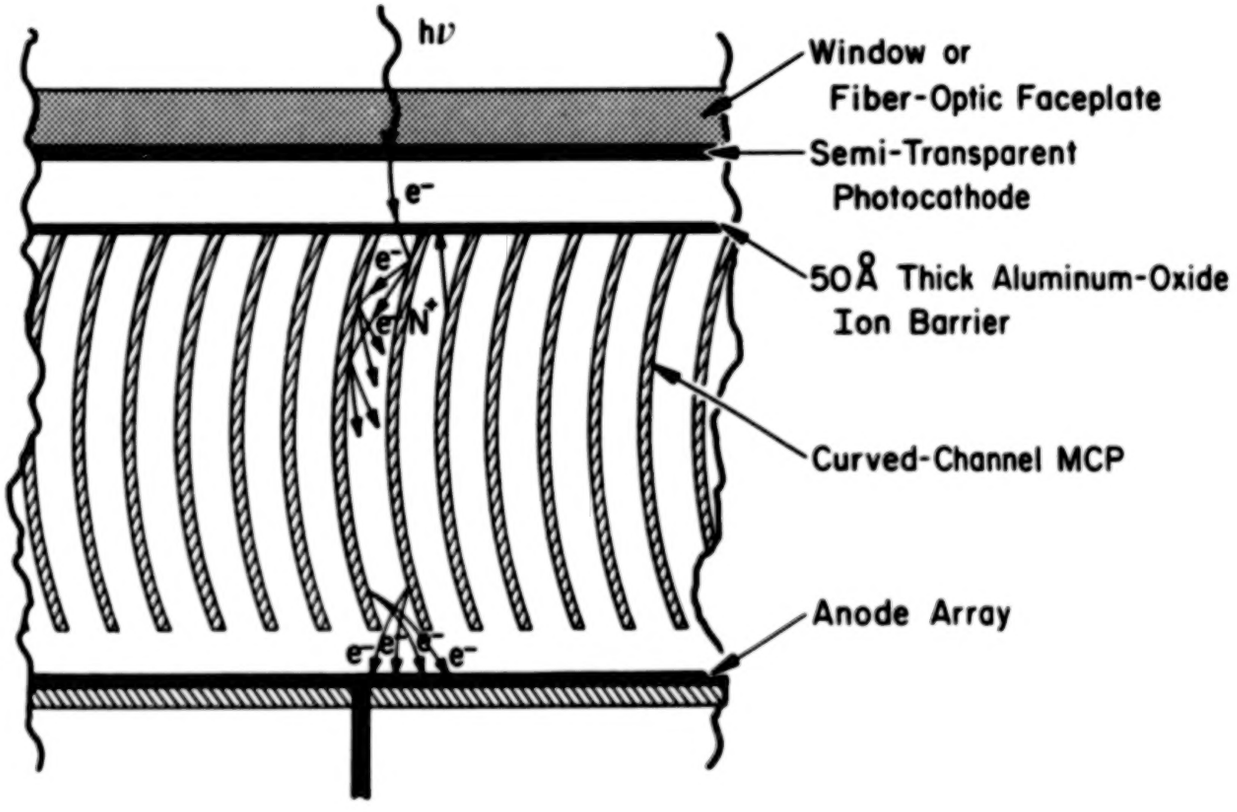


Figure 3. Schematic of visible-light MAMA detector tube utilizing a proximity-focused, semi-transparent photocathode.

The MAMAs have a number of unique characteristics. The remainder of this section deals with those characteristics which affect the photometric performance of the detectors.

(a) The Pulse-Height Distribution

The high gain and narrow output pulse-height distribution of the curved-channel MCP provide a photometrically stable response without stringent requirements on the stability of either the high-voltage power supplies or of the discriminator threshold (see, for example, Timothy 1981).

Because of the effects of space-charge saturation within the channel, the curved-channel MCP produces a sharply peaked output pulse-height distribution (see Figure 4) in marked contrast to the almost exponential output pulse-height distributions produced by conventional discrete-dynode photomultiplier tubes.

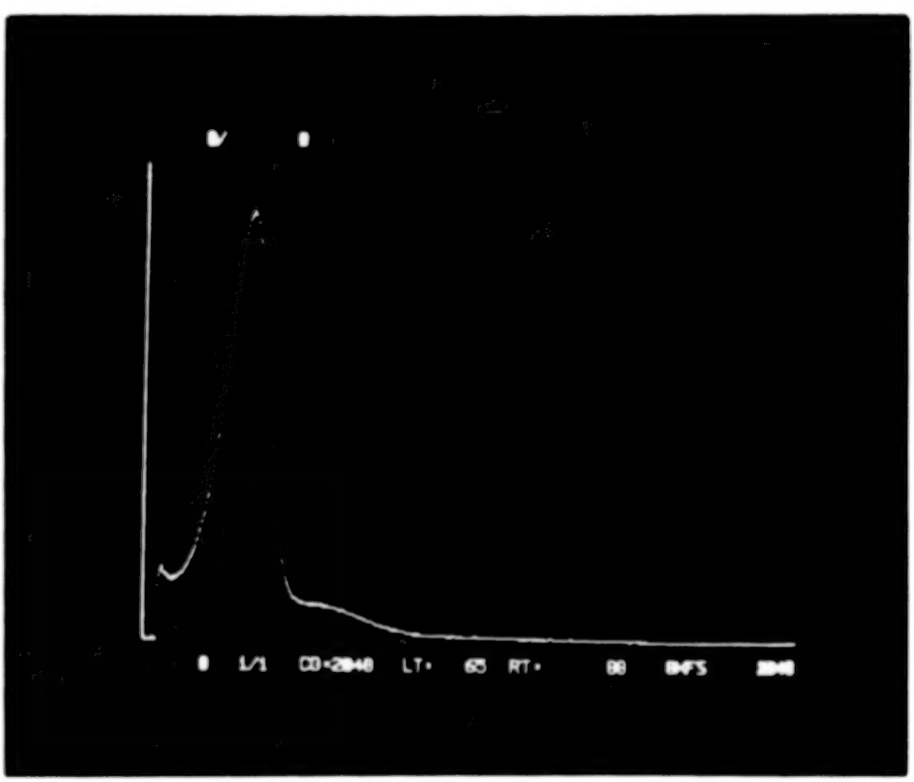


Figure 4. Output pulse-height distribution for a curved-channel MCP with 25-micron-diameter channels. Modal gain is 1.3×10^6 electrons pulse^{-1}, and the resolution (FWHM) of the distribution is 40%.

This sharply peaked output pulse-height distribution provides a number of advantages when MAMA detectors are used for photometric measurements. First, if the discriminator threshold is set at a level well below the peak of the distribution, the requirements on the stability of both the discriminator threshold and the applied high-voltage can be relaxed since small threshold changes or small gain changes caused by slight changes in the applied high voltage will cause essentially no variation in the detective quantum efficiency.

As the output count rate of the MAMA detector is increased, the MCP gain will decrease because of the reduction of the available charge on the channel walls. For a curved-channel MCP with a sharply peaked output pulse-height distribution, significant changes in the detective quantum efficiency will not be observed until the count rate has reached the level at which the peak of the distribution falls to the level of the discriminator threshold. The rate at which the gain decreases is determined by the time constant of the channel and is controlled principally by the channel conductivity. For typical MCPs with conductivities of the order of 100 MΩ, output count rates in excess of 10^5 counts mm^2 s^{-1} (random) can be obtained with a loss of detective quantum efficiency of less than 10%.

Since the gain of the MCP at low count rates is not dependent on the conductivity of the channel but only on the physical dimensions, temperature changes will not affect the gain or the detective quantum efficiency. Temperature changes will, however, affect the conductivity of the MCP and therefore the dynamic range, and stabilized high-voltage power supplies will be needed to maintain a constant photometric response.

When properly conditioned by the appropriate bake and scrub procedures, a high-gain, curved-channel MCP displays excellent gain stability as a function of time provided that it is operated in a hydrocarbon-free, high-vacuum

environment. Studies of long-term gain stability are currently in progress, and accurate measurements of the long-term photometric stability of MAMA detector tubes are expected to be underway at Stanford University by the end of this year.

(b) Zero Readout Noise

Since the MAMA detectors operate with zero readout noise, the system is limited by photon statistics noise. The signal-to-noise ratio of the photometric data is given by:

$$\rho = \frac{(N - D)\cdot t}{[(N + D)t]^{1/2}}$$

where N = total count rate (signal + dark)

D = dark count rate

t = integration time.

(c) High Detective Quantum Efficiency

Typical quantum efficiencies for ultraviolet and visible-light photocathodes are shown in Figure 5. As for conventional photomultiplier tubes, the MAMA detectors have pulse-counting detective quantum efficiencies that are typically 70 to 80% of the photocathode quantum efficiencies. Typical dark count rates are of the order of 0.1 counts mm^{-2} s^{-1} at room temperature for the ultraviolet Cs_2Te photocathode and in the range 0.2 to 1 counts mm^{-2} s^{-1} at temperatures from 0°C to -30°C for the visible-light bialkali and trialkali photocathodes.

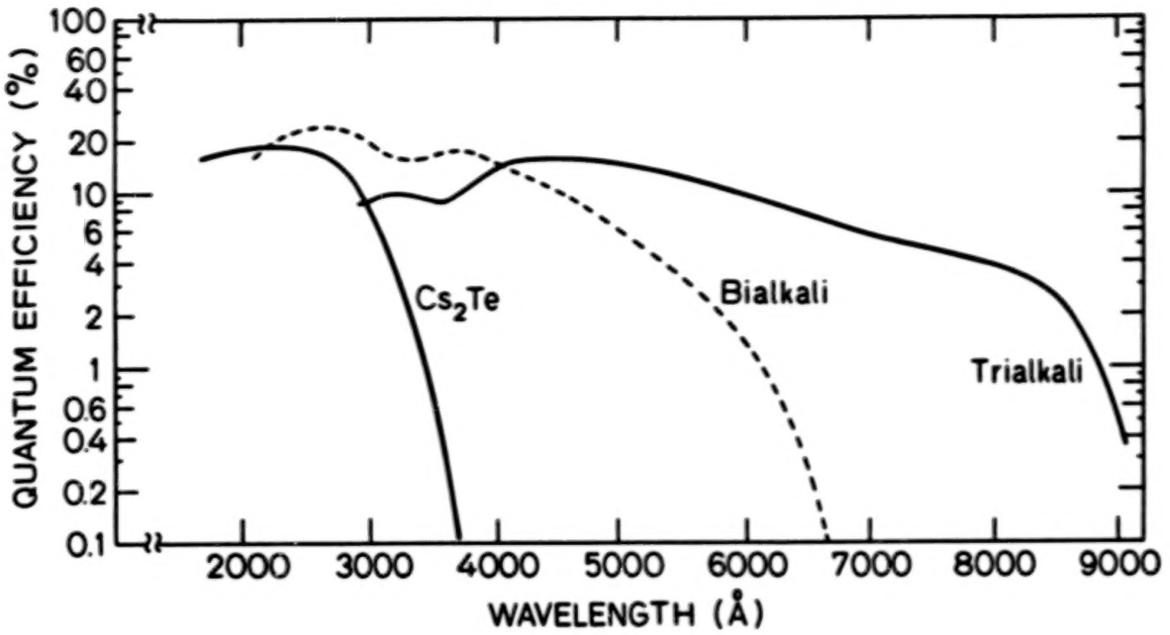


Figure 5. Quantum efficiencies of ultraviolet and visible-light MAMA detector tubes with semi-transparent photocathodes.

(d) Dynamic Range

For photometry of extended objects (surface brightness measurements), the MAMAs have a high dynamic range. A 10% coincidence loss is obtained at a signal level in excess of 10^5 counts mm^{-2} s^{-1}, and the present electronics can handle up to 10^6 counts s^{-1} (random) with a 10% coincidence loss.

For point sources, the coincidence loss sets a limit on how bright a star can be measured. With a 1 meter telescope, a bandpass of 300 Å, and adopting an overall efficiency of 10%, a star of visual magnitude 14 will, if imaged onto a single pixel, produce a count rate from that pixel of 6.5×10^2 counts s^{-1}, which corresponds to 10% coincidence loss. In practice, the star image will be spread over an area several-pixels by several-pixels (depending on the image scale). In addition, substantially narrower bandpasses may be dictated by other considerations; and, finally, there are prospects for improving the

dynamic range by several orders of magnitude in the near future. In the meantime, it is clear that the MAMA detector system can be used on bright stars only if some means of reducing the count rate is employed. Another way of looking at this limit is to note that the count rate of a star is limited to 6.5×10^2 times the number of pixels the star image is spread over. For an image scale of about 10 arc sec mm^{-1} and a seeing disk of 1 arc sec in diameter ($\sim 4 \times 4$ pixels), this means that the maximum rate is about 10^4 counts s^{-1}. A statistical accuracy of 0.1% can then be achieved in several minutes of integration. Note that 100 such stars could be imaged simultaneously on the array without exceeding the total count rate of 10^6 counts s^{-1}, corresponding to a 10% coincidence loss.

(2) CHARGE-INJECTION DEVICES

The Charge-Injection Device (CID) differs from the Charge-Coupled Device (CCD) in a number of important aspects. In the CID, the accumulated charge is not transferred serially out of the array but is shifted between two adjacent solid-state capacitors as shown schematically in Figs. 6a and 6b. Charge transfer only occurs between an individual pair of electrodes, and not all charge packets are moved to the same common output. Since the charge is stored for as long as at least one of the electrodes is biased on, this is a non-destructive technique and the readout noise of the device can be significantly reduced by reading out the signal a large number of times. To clear the device, the stored charge is injected into the substrate which provides an alternative readout technique (Fig. 6c). Comparisons of the CID with the CCD have been given earlier by Barbe (1975) and by Burke and Michon (1976).

To date, relatively little work has been done with CIDs in comparison with the large amount of evaluation that has been undertaken for the CCDs.

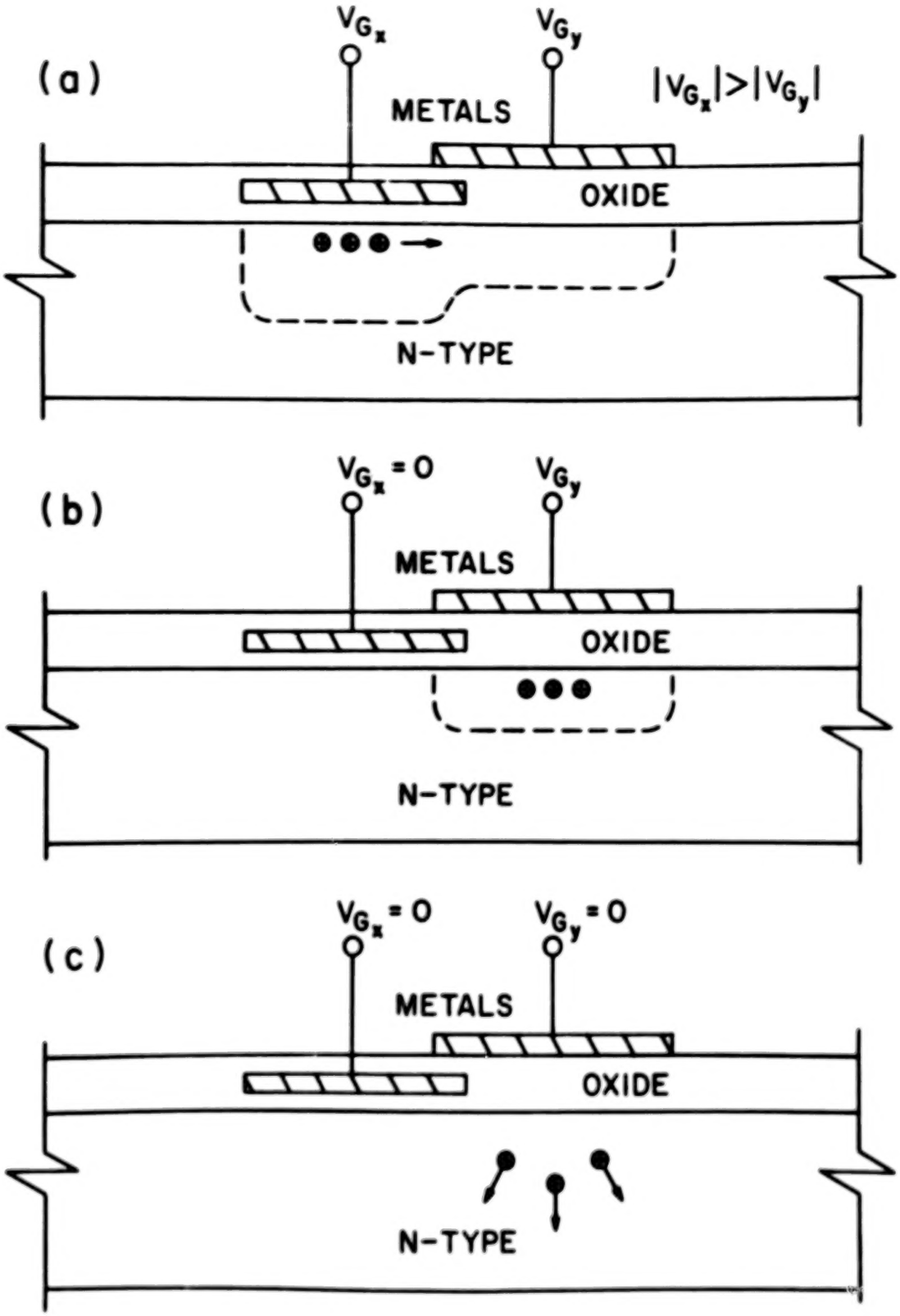


Figure 6. Schematic of the basic cell of a CID area array.
a. Charge storage.
b. Charge transfer.
c. Charge injection (from Barbe 1975).

Principally, the major work on CIDs has been carried out at GE.[1] For example, an experimental chip with a format of (128 x 128)-pixels has been evaluated for use as a stellar tracking device (Burke, et al. 1979). This chip has thin polysilicon electrodes and a peak quantum efficiency in excess of 50% at 7000 Å. Readout noise levels are about 40 electrons rms for multiple readouts.

CIDs with low readout noise levels could be extremely attractive as photometric detectors at optical wavelengths for a number of reasons. First, the CID has a full-well capacity greater than that of the best CCDs currently available. Second, the readout technique eliminates the problems associated with charge transfer inefficiencies at low charge levels that are observed in some of the CCDs, and spurious charge injection should not be a problem. Third, the non-destructive, multiple-readout capability allows cosmic ray events to be detected during a long integration and rejected from the final data set. Finally, the x-y addressable readout configuration eliminates the limitations of the line-sequential readout system of the CCD and permits subsets of an array to be read out under external control.

REFERENCES

Barbe, D. F. 1975, Proc. I.E.E.E. 63, 38.

Burke, H. K., and Michon, G. J. 1976, I.E.E.E. Trans. on Electron Devices, ED-23, 198.

Burke, H. K. , Michon, G. J., Tomlinson, H. W., Vogelsong, T. L., Grafinger, A., and Wilson, R. 1979, NASA Marshall Space Flight Center, Final Report NAS8-32801.

Timothy, J. G. 1981, Rev. Sci. Instrum. 52, 1131.

---

1 General Electric, Schenectady, NY 12305

Timothy, J. G., and Bybee, R. L. 1981, SPIE Shuttle Pointing of Electro-Optical Experiments 265, 93.

Timothy, J. G., and Bybee, R. L. 1983, Proceedings of AIAA 21st Aerospace Sciences Meeting (Reno, NV).

NOTE: This material has been abstracted in part from a paper, "Optical Detectors for Spectroscopy," which appeared in Publications of the Astronomical Society of the Pacific 95, p. 810, 1983.

Jeff Scargle of NASA Ames Research Center assisted in the preparation of this article.

IV. OPTICAL PROBLEMS

Filter and Passband Problems

Andrew T. Young
San Diego State University

INTRODUCTION

For both instrumental and astrophysical reasons, all precise photometric measurements of stars must be made with restricted optical bandwidth. If two measurements are to be compared precisely, they must have precisely the same passband. To give some notion of the reproducibility required, I note that the range of typical stellar color indices for the usual 1000-A color baseline is about 2 magnitudes, so that an error of 1 A in effective wavelength can already introduce errors of .002 mag or so. Furthermore, the band width must be maintained to the precision required in the photometry. For the intermediate-width passbands popular today, which are about 250 A, this means the band width must be constant to 1/4 A if we want .001 magnitude precision.

GLASS FILTERS

This precision is difficult to achieve instrumentally. For sharp-cutoff glass filters, which (like most solutions and other amorphous materials) obey Urbach's rule, the thermal shifts in filter passbands are typically 1 A/degree in the blue, and vary as the square of the wavelength (Young, 1974, p. 106). Because this shift occurs mainly at the blue side of the filter passband (cf. Young, 1967), this is also the order of magnitude of the change in band width; so apparent filter transmissions change by several tenths of 1%/degree for the band widths of a few hundred Angstroms that are achievable with glass filters.

Absorbing filters with a transmission passband (e.g., blue glasses) usually have an Urbach edge on the short-wavelength side of the passband, and an absorption band (usually due to transition metal ions) with a Gaussian profile on the long side. This absorption broadens with increasing temperature, and shifts slightly as well, due to changing solvent shifts as the glass matrix expands. The combined effect is small on the blue side of the absorption (i.e., the red side of the passband), and large on the red side of the absorption. Again, the effects are some small multiple of kT in photon energy at the band edge, or a considerable fraction of 1% per degree in the visible. This value increases strongly for thicker filters and narrower passbands, in which there is appreciable absorption at the transmission peak.

Thus, glass filters must be thermally stabilized to one or two tenths of a degree to yield .001 mag reliably. This is near the limit of what is technically practical, because glass has so low a thermal conductivity that a filter's temperature is determined primarily by exchange of thermal infrared radiation with its environment, not by the temperature of its holder.

To be sure, some of this variation cancels out in differential photometry.

But because the filters change much more at the blue side of the passband than at the red edge, the effects are appreciably more severe for blue than for red stars. Thus, McCullough (1964) found the apparent V magnitudes of blue stars to change relative to red ones by 0.001 mag/degree per unit of color-index difference; this value agrees with numerical integrations based on filter and photomultiplier temperature coefficients measured in the laboratory (Young, 1967). Similarly, color zero-points change by typically 0.002 mag per degree if glass filters are used.

In fact, as the instrumental temperature changes, so does the instrumental response. Observations made at different temperatures are actually made by different instruments, so one must confront the problem of transforming from one instrumental system to another (see Young, 1974, pp. 180-192; Manfroid, 1984). This is discussed below.

INTERFERENCE FILTERS

If we use interference filters, the thermal effects are about 5 times smaller, and are due to the thermal expansions of the layers and the thermal dependences of the refractive indices (and hence the optical thicknesses) of the materials. Unfortunately, these filters have severe problems with long-term stability. The changes can be tens of per cent in a few weeks in unfavorable cases; in the best cases, lifetimes of a few years are expected. This is largely due to the infiltration of water vapor, which changes the thicknesses and refractive indices of the dielectric layers, thus shifting and degrading the passband. These long-term changes in spectral response, and the eventual replacement of old filters, again introduce the problem of transforming data taken at one time to the system of those taken at another. If the characteristic time scale is 3 years, the changes are significant at the 10^{-3} mag level from one night to the next, and at the 10^{-4} mag level in 2 or 3 hours.

It should be possible to prolong the life of interference filters by keeping them in a dry environment, but the method of maintaining dry air in the photometer has yet to be worked out. One would have to seal the instrument with a window, which might cause problems with inter-reflections. The photometer could be filled with dry nitrogen, or evacuated, but extremely tight seals would be required to prevent water vapor from diffusing in. Also, one would have to use oil-pumped nitrogen, which requires careful filtering to remove microscopic oil droplets that could contaminate optical surfaces. The problems are difficult, but perhaps not insoluble. While the humidity problem would be eliminated by operation in space, there might still be problems due to recrystallization in filter layers, due to infiltration of volatile contaminants other than water vapor, and due to radiation damage. More stable filters appear possible (Martin et al., 1984), but are not yet available.

Another problem is the angular dependence of interference filters. Because the large-scale spatial nonuniformity of these filters amounts to several per cent (Schaefer and Eckerle, 1984), one must image the pupil on the filters before re-imaging it on the detector. But then when the star image moves (due to seeing or poor telescope tracking), the angle of incidence changes. Apart from a factor of order unity due to the effective refractive index of the spacer layers, the

fractional wavelength shift is half the square of the incidence angle. If we image a one-meter telescope down to 1 cm on the filter, a one arcsec image movement on the sky is 100 arcsec at the filter, or 5×10^{-4} radians. The corresponding passband shift is about 10^{-7} of the wavelength per arcsec (squared) of image motion. Typical image motion associated with "seeing" is a few arcsec, corresponding to fractional passband shifts of 10^{-6} on-axis.

Unfortunately, several per cent of the light in a star image lies more than 10 arcsec from the optical axis, where the angular effects are 100 times larger (as they increase with the square of the off-axis angle). The observed inverse-square drop of intensity with distance from the image center means that successive zones of the image have greater effects on the passband centroid. Thus the effect of bad seeing or tracking is to produce a long blue wing on the effective filter passband. The larger the focal-plane aperture we must use, the worse this effect becomes. With Don Kurtz's 1-arcmin diaphragms, centroid shifts of several tenths of an angstrom can be expected, and will change with seeing and guiding errors.

From the discussion above, we see that photometry much more precise than .001 mag is not likely with any kind of filters.

BAND WIDTH PROBLEMS

Another major problem is the correct analytic representation of bandwidth effects (Blanco, 1957) in data reduction. The usual technique (cf. King, 1952; Young, 1974, pp. 154-161) is to include a linear color term in the transformation from observed to extra-atmospheric magnitudes. This is simply a first-order transformation from the (instrument + atmosphere) system to the (instrument alone) system. As I have pointed out (Young, 1974, p. 158), the correct color to use is neither the extra-atmospheric color nor the observed color, but their mean. This automatically produces the quadratic airmass term that is usually used to represent the Forbes (1842) effect.

However, numerical integrations (Young, 1974, p. 159) agree with direct measurements (Hardie, 1966) in showing that the curvature of the Bouguer curve decreases with increasing airmass, while the simple theory cited above predicts that it should increase. Though I mentioned that this is due to the increasing monochromaticity of the transmitted light as more and more of the strongly-extincted rays are removed, I failed to point out that this is not included in the theoretical development. I believe the main weakness lies in the approximation, in King's classical paper, of $\ln(1+x)$ by x, even though "x" is not small at large airmass. However, the whole theory needs re-examination and fuller development, as it does not agree with the numerical integrations better than several thousandths of a magnitude, over the range of airmass that must be used if the extinction is to be determined with reasonable accuracy.

It is sometimes supposed that these bandwidth effects are "negligible" in intermediate-bandwidth systems like uvby. However, the size of the color term in the transformation from inside to outside the atmosphere is proportional (in King's approximation) to the square of the bandwidth. As bandwidths are only a factor of 4 or 5 smaller for uvby than in UBV, the color terms should be about 20

times smaller. Typical color-term coefficients in UBV are 0.03 mag/airmass, so in uvby they should be about 0.0015 mag/airmass in v and b, and several times larger in the wider u band.

If the bandwidths were reduced much more, the increased integration times required to reach a precision of 10^{-3} or 10^{-4} mag would become a problem, if a reasonable number of stars are to be within reach. Indeed, as the airmass typically changes by 0.02/min for the extinction stars at 2.5 air masses, and the brightness is continuously changing (by, say, 0.004 mag/min in B, and about twice as much in U), the definition of a proper "mean" time and air mass to assign to integrations longer than a minute or two becomes a problem between the 10^{-3} and 10^{-4} mag levels, even for the bright stars limited only by scintillation noise. (Scintillation is about 0.003 mag at airmass 2.5 for a 10-sec integration at a 1-meter telescope, so about 100 sec are required to reach 0.001 mag precision in measuring the extinction.)

The linear color-dependent term in the extinction that can be derived from King's theory is an approximate way to correct the atmospherically-reddened observations to their extra-atmospheric effective wavelengths. The approximation uses some color index instead of the actual gradient of the stellar monochromatic magnitude. As mentioned above, the mean of the observed and extra-atmospheric colors should be used.

SPECTRAL SAMPLING

However, none of these colors is strictly correct in any existing photometric system. All such systems violate the sampling theorem, and the use of their colors introduces aliasing errors (Young, 1974; p. 188). Thus, in the UBV system, two stars can have the same (b-y) color, but their effective (b-y) extinction coefficients can differ by 0.01 mag/airmass (see Young, 1974, Fig. 14, p. 158; cf. Mandwewala, 1976, Fig. 6). Because all existing systems are undersampled, they fail to provide information (such as local spectral slopes) needed for accurate data reduction. A fully transformable system is needed just to relate observations of different objects observed at different airmasses with the "same" instrument, because (as mentioned above) the atmosphere must be considered part of the instrumental system, which is not really the same at different zenith distances.

Manfroid (1984) has shown that the gradients required to do this linear interpolation can be measured, so that the interpolation can be done much more accurately, if an additional band is placed near each existing one. This would provide spectral gradient data at each existing band, but without adding the large number of filters that would be needed for complete spectral sampling.

A fully sampled system would have an additional advantage: the same linear transformations would apply to all stars, regardless of luminosity, age, metallicity, or interstellar or atmospheric reddening. This would eliminate the paradoxical need for "standards" in regions of the observational parameter space inhabited only by variable objects.

The number of bands required for complete sampling can be reduced if they

have smooth, symmetrical profiles whose Fourier transforms die out rapidly (e.g., Gaussian or cosine-squared profiles). These can be approximated with existing filter-design methods (Wright et al., 1969; Falletti et al., 1982), but might be approached even more closely with some non-filter method, perhaps using a spectrometer.

REFERENCES

V. M. Blanco, Ap.J. 125, 209 (1957).

V. Falletti, A. Premoli, and M. L. Rastello, Appl.Opt. 21, 4345 (1982).

J. D. Forbes, Phil.Trans. 132, 225 (1842).

R. H. Hardie, in IAU Symp. 24 (Academic, 1966) p. 243.

I. R. King, Astron.J. 57, 253 (1952).

N. J. Mandwewala, Pub. Obs. Cenève Ser. A, Fasc. 82 (Arch. Sc. Genève 29), 119 (1976).

J. Manfroid, "On photometric standards and color transformation," IAU Symposium III (1984).,

P. Martin, W. Pawlewicz, D. Coult, and J. Jones, Appl.Opt. 23, 1307 (1984).

J. R. McCullough, A.J. 69, 251 (1964).

A. R. Schaefer and K. L. Eckerle, Appl.Opt. 23, 150 (1984).

H. Wright, C. L. Sanders, and D. Gignac, Appl.Opt. 8, 2449 (1969).

A. T. Young, M.N. 135, 175 (1967).

A. T. Young, In Methods of Experimental Physics, Vol. 12, Part A: Astrophysics, Optical and Infrared. (Academic Press, 1974) Chapters 1-3.

Light Scattered from Polished Optical Surfaces: Wings of the Point Spread Function

C. E. KenKnight
University of Arizona
Tucson AZ 85721

I. Abstract

Random figure errors from the polishing process plus particles on the main mirrors in a telescope cause an extended point spread function (PSF) declining approximately as the inverse square of the sine of the angle from a star from about 100 μrad to a right angle. The decline in at least one case, and probably in general, proceeds as the inverse cube at smaller angles where the usual focal plane aperture radius is chosen. The photometric error due to misalignment by one Airy ring spacing with an aperture of n rings depends on the net variance in the figure. It is approximately $60/(n+1)^3$ when using the data of Kormendy (1973). A typical value is 6×10^{-5} per ring of misalignment with n = 100 rings (100 μrad hole at .5 μm with a 1 m telescope). The encircled power may be modulated on a time scale of hours by parts per thousand in a wavelength-dependent manner due to relative humidity effects on mirror dust. The scattering according to an inverse power law is due to a random walk in aberration height caused by a multitude of facets and slope errors left by the polishing process. A deviation from such a law at grazing emergence may permit monitoring the dust effects.

II. Introduction

We are concerned here about the extended PSF caused by surface errors of telescopes used for photometry. Misalignment of a star in a focal plane stop will cause systematic error in the photometry. Our task will be an estimate of that error.

Understanding the light distribution near the center of the PSF for a real telescope in the 1 to 5-m class is not easy. Reflection of light from the surface introduces wavefront aberrations with a spectrum of roughness from half the diameter D down to scratches or other debris unresolvable by the wavelength used. If the total rms measure of that roughness is σ in radians, the power in the central peak and surrounding Airy rings is decreased (e.g., Porteus, 1963) by $\exp(-\sigma^2)$. Aberrations larger than about $\pi/2$ call for selective

polishing, so a σ of about 1 radian is usual. Thus two thirds of the light is redistributed in unwanted places. The surface does not qualify as either very smooth or rough, the extremes for which theories tend to be designed (Elson & Bennett, 1979a). My computer simulations and the experience of others with final acceptance testing (private communication from R. E. Parks & A. Meinel, 1984) show that no Airy rings can be discerned for $\sigma > 1$. The aberrations redistribute the power irregularly so that the rings are obliterated by a speckle pattern.

At wide angles Θ relative to the (normal) optical axis the effective grating that is scattering the light has a period

$$d = \lambda / \sin \Theta \qquad (1)$$

much smaller than the aperture diameter D. The focal plane light distribution is there caused by $N = 2\ D/d$ strips of grating across the pupil and usually some subdivision of those strips into independent areas, a huge number of scattering regions. The distribution of light is then a fully developed speckle pattern (Goodman, 1975). Experiment shows (Redman & Shirley,1938; Zanstra, 1943; Pierce, 1954; Gehrels et al., 1964; King, 1971; Piccirillo, 1973; Kormendy, 1973; Harvey, 1976; Thomas, 1980; Wang & Wolfe, 1983) that the power falls off there roughly as $1/\sin^2 \Theta$ in angle and $1/\lambda$ in wavelength. For limited regions of angle the power may fall off a little less rapidly with angle where dust or patchy films make a contribution (Thomas, 1980). The data of Kormendy (1973), shown in Figure 1, have a region of slope -2 becoming slope -1.7. At angles below 100 μrad Kormendy found that scattered light power fell off at the Θ^{-3} rate characteristic of edge diffraction, but the power is an order of magnitude too large for edge diffraction. These data for blue light using the 48-inch Palomar Schmidt telescope are alone in this critical angular range for our problem, so we must use them. We will need to consider whether his -3 slope should be as universal a phenomenon as the -2 slope at larger angles is known to be. At these smaller angles, however, the number of patches contributing to the speckle pattern becomes small; the statistics problem becomes difficult, even hazardous. We will resort to a symmetry argument to affirm that the -3 slope should be generally present. We discuss also the physical basis for these power laws.

III. Photometric Error for Misalignment

The familiar Airy ring pattern has an irradiance due to edge diffraction alone of the form

$$B(r) = [2\ J_1(r)/r]^2 \qquad (2)$$

Using the asymptotic form for the Bessel function $J_1(r)$ at large argument

$$J_1^2(r) \cong (2/\pi r)\cos\phi,\ \phi = r - 3\pi/4. \quad (3)$$

Counting the Airy ring maxima by n starting with n = 0 at the center we have

$$r_n \cong \pi(n + 3/4) \quad (4)$$

The angular units counted by n are

$$\theta_o = \lambda / D. \quad (5)$$

Averaging the irradiance between the dark rings gives

$$B(r_n) \cong 4\pi^{-4}(n + 3/4)^{-3} \quad (6)$$

A central obscuration of size tD causes a waveform in angle (Young, 1974) that is a beat pattern of two edge diffractions and increases the mean irradiance by a factor $(1+t)/(1-t^2)^2$, which is 2.7 at t = 0.5.

The experimental star profiles are given as radiances. Kormendy (1973) gave a radiance in magnitudes per arcsecond squared that translates to 2.7×10^{-6} μrad^{-2} at 29 μrad = 79 rings for D = 1.22 m and θ_o = 0.37 μrad. For that number of rings (6) gives $B(r_n) = 8.1 \times 10^{-8}$. To change this to a radiance we use the fraction of the theoretical edge-diffracted power inside the first dark ring at 1.22 θ_o, $1-J_0^2(3.833) = 0.838$, and the solid angle $\pi(1.22\ \theta_o)^2$ to obtain a central radiance unit for the unobstructed telescope without figure errors

$$b = 0.179\ \theta_o^{-2}. \quad (7)$$

Then one edge diffraction would give a radiance

$$bB(r_n) = 1.07 \times 10^{-7}\ \mu rad^{-2}, \quad (8)$$

which is K = 25 times smaller than the Kormendy radiance. Part of the discrepancy would be removed if we had considered diffraction by the central obscuration of the Schmidt telescope. We should have increased both $B(r_n)$ as stated above and b, because the first dark ring occurs slightly before 1.22 θ_o. A factor of 8 to 10 discrepancy remains, however.

Suppose, then, that a focal plane aperture is misaligned by one ring. The discrepant area on one side is nearly a sine wave of solid angle πr_n and the power change per ring is

$$\Delta P = \pi^2(n + 3/4)(B_{n-1} - B_{n+1})K$$
$$\cong 24\ K\ \pi^{-2}(n+1)^{-3} \quad (9)$$

A common aperture radius is 130 rings for which ΔP is 2.7×10^{-5} per ring of misalignment. My computer studies on seeing-perturbed wavefronts show that instantaneous alignment to less than 1 ring is not possible (or meaningful) because central measures of the asymmetric PSF do not agree at that level. Setting the star in the aperture by eye assures error of 10 to 20 rings, which implicates misalignment as a substantial photometric error source for the future systems. Recall also that the eye is nearly a V filter and that the atmosphere dispersion without the use of a compensating Risley prism will increase the

errors for passbands other than the V filter.

IV. Scattered Light Power and the Kormendy Discrepancy

We noted that in final testing of large mirrors the seeing of any Airy ring is rare. But suppose to the contrary that the Kormendy curve did bend over and become constant inside the position for dark ring 3 at a radiance equal to the mean radiance for the edge diffraction ring 2 (no central obscuration),

$$C_3 = bB(r_2) = 3.5 \times 10^{-4}\ \theta_o^{-2} \qquad (10)$$

Inside 2 dark rings we would accumulate a power due to surface errors

$$P_3 = \pi C_3\ (2.2\ \theta_o)^2 = 5.3 \times 10^{-3} \qquad (11)$$

Beyond 2 dark rings the net power for the Kormendy curve would then be roughly $J_0^2(r_2) = 0.090$, since it mimics the angular dependence of edge diffraction. The total power would then be only 0.095 and σ would be 0.31 radian = 0.05 wave. But this is improbably perfect for a 48-inch telescope. Rather a σ of about 1 radian is to be expected, so it is self-consistent that the Kormendy curve is 8 to 10 times higher than edge diffraction all the way to 2 rings.

Beyond 150 μrad the θ^{-2} descent starts at about 1.7×10^{-8} μrad^{-2}. A $\sin^{-2}\theta$ curve would contribute an additional power

$$P_2 = 2\pi(1.7)\ 10^{-8}\ (150)^2 \ln \cot(75\ \mu\text{rad}) = 0.023 \qquad (12)$$

which is small in comparison to the fraction 0.7 or so of the total power that is represented by the θ^{-3} region below 150 μrad. The lesson is that the more slowly descending θ^{-2} redistribution of light is caused by only a few percent to the aberration variance. Nevertheless, the slow θ^{-2} descent assures dominance of the mechanism at large angles.

V. Surface Aberrations as a Random Walk

Keshner (1982) in a seminal paper on 1/f noise showed that such noise arises in a system where the time derivative of some quantity has a white spectrum. In his example the quantity was charge at a given point on an infinitely long resistor-capacitor transmission line. Input to the line was a white current source. Such a system has an infinitely long memory and the variance of the charge is infinite, but would require an observation time interval of infinite length for its measurement. It is an example of a nonstationary random process. Another example is the Brownian motion of a particle or any such Markov process. Keshner showed that on any finite observation interval short compared to the time elapsed since the process began, such systems can be approximated as a stationary random process for which an autocorrelation function (ACF) is defined. The Fourier transform of

that ACF is the power spectrum which is commonly compared with experiment. A helpful way to think of the power spectrum is that it is a kick-sorter by frequency of the variance. The normalization will therefore be arranged so that a sum over all frequencies gives all the variance. When so normalized, we will refer to it as a power spectral density function.

In the context of rough surfaces the time derivative becomes a spatial derivative, the frequency a spatial frequency, the quantity of interest becomes the height of a rough surface relative to a mean (and finite) optical surface, and the height is the summation over facets having a range of sizes and slopes. Using light in reflection as a probe of the height z, the regime $4\pi z/\lambda << 1$ allows a one-to-one mapping from the power spectrum of z(x,y) to the power spectrum of light scattered away from the specular direction (e.g., Church, 1983, eq. 19). The power spectrum of z is expressed in terms of spatial frequencies, say, p and q. The light spectrum is described (Church et al., 1977) as in Figure 2 in terms of angle of incidence Ψ and the angles of emergence θ and ϕ such that for $k = 2\pi/\lambda$

$$p = k(\sin\theta\cos\phi - \sin\Psi), \quad q = k\sin\theta\sin\phi \qquad (13)$$

In the plane of incidence we have vanishing ϕ, or

$$p = k(\sin\theta - \sin\Psi), \quad q = 0. \qquad (14)$$

Experiments on clean polished surfaces (Harvey, 1976; Thomas, 1980; Wang & Wolfe, 1983,1984) showed light power

$$P \propto I\,A\,\Omega\,f\,k\,\sigma^2\,p^{-2} \qquad (15)$$

where I is the incident irradiance, A the illuminated area, Ω the detector solid angle seen at the sample, and the factor

$$f = \cos^2\theta\cos\Psi\,Q \qquad (16)$$

contains angular factors of the order of unity for the present purpose plus a factor Q containing the dependences on incident and final states of polarization and the polarizing properties of the bulk material. Usually the experiments use incident and final electric field vectors perpendicular to the scattering plane (s polarization) to avoid exciting ephemeral surface waves that can extract small (but difficult to predict) power from the system.

The power spectrum of z must not continue to small p as p^{-2} indefinitely. Apart from a possible additive function (such as the p^{-3} region in the Kormendy data), the spectrum must terminate, for example, as the Lorentzian function $2\pi^{-1}g/(g^2 + p^2)$, for which the normalized ACF is $\exp(-|gx|)$ in one dimension. An exponential ACF implies a Markov process (e.g., Doob, 1953, pg. 233). Therefore we are forced to consider that at least part of the surface deviation z strongly resembles a random walk driven by facets of various slopes and a continuous size range down to at least the wavelength of light.

The general character of the problem of a power-law decline of the

Fourier transform is revealed by a theorem on discontinuities given by Bracewell (1978). When at high frequency p the spectrum of a function decays as slowly as $p^{-(n+1)}$, then n is the lowest derivative of the function that has at least one discontinuity. When experiment shows that scattered light power falls as p^{-2}, we can conclude that the ACF of the fields in the radiating aperture has at least one discontinuity in the first derivative. In particular, the exponential ACF has such a discontinuity at zero lag. The underlying random process may be continuous in a finite interval, but it may have a discontinuous derivative anywhere (Dainty, 1984). If the power falls off as p^{-2}, then the modulus falls off as p^{-1}. Since $n = 0$ for that case, there is at least one discontinuity in the function itself, proportional to the unit step function studied by Papoulis (1962, pg. 260). My computer models (section VI) of a random walk have such spectra but no discontinuity in the studied interval. We have to ask where that discontinuity is located. The answer is that the one discontinuity is at infinity. The change in the random-walk function for an infinitely long interval diverges.

If a discontinuity at infinity governs the power-law decay of the infinite-interval Fourier transform, then how is it that a finite interval can reveal the behavior at infinity? The answer can be seen from a discussion by Sayles & Thomas (1978). They collected data from a wide variety of macroscopic surfaces having power spectra that decay as p^{-2} at high spatial frequency p. Their examples ranged from the raceways of bearings through ship hull plates to planetary surfaces. In those cases for which enough data were available, the variance per unit length increased linearly with the length of the subinterval used to calculate the variance. For a stationary process the variance per unit length is constant, independent of the subinterval used. This quasi-local property reveals the behavior at infinity.

On macroscopic surfaces there is no difficulty in making measures along a straight line or inside a linear strip of arbitrarily small width. The light scattering from microscopic roughness usually involves a two-dimensional surface. The new features of two dimensions are reviewed in section VII.

An exponential ACF for polished samples was demonstrated by Eastman & Baumeister (1974), Bennett (1976), and Elson & Bennett (1979b) using interferometers and a scanning stylus, direct methods. An exponential ACF was reported for sanded glass by Rezette (1979) using a two-exposure holographic method not included in the summary of methods given by Church (1983). The method is recommendable for the present argument because it does not omit any of the scattered radiation at the small angles where most of the scattered radiation occurs; yet the ACF was exponential. On the other hand, it seems suited only

to small samples since the sample must be translated in an accurately known manner. It also does not have the flexibility of studies with a scatterometer (e.g., Stover et al., 1983) of roughness in various spatial frequency bands. I insert this remark in deference to the reality that surfaces in experiment are not simple. The ACF is not a pure exponential due to a random walk plus a delta function at zero lag due to unresolvable edges of facets. The power spectrum is not a perfect Lorentzian above a white background. Idealizations of models are useful for tutorial purposes but little else.

VI. Computer Models of Surface Roughness

VI.A. Markov Chains

When observing spatial frequency p, the surface is effectively multiplied by the grating exp(ipx). Regions of width π/p are averaged in the height z and summed to give a net electric vector in a Kirchoff diffraction integral (KenKnight, 1977). The number of strips that are to be added in the angular region where the θ^{-2} scattering holds is so large that the probability distribution for the scattered field is a complex Gaussian, whatever the distribution for z might have been. A more precise statement of the inaccessibility of the z distribution given small z and fixed p was made by Beckmann (1973), "To a very close approximation, therefore, all slightly rough surfaces scatter like a normal surface of the same slight roughness."

When changing the spatial frequency p, the nature of transverse correlation on the surface is probed. I made ensembles of one dimensional surfaces in a computer as follows: select a facet size s from the square of a Gaussian number generator (i.e., from a chi-squared distribution with 1 degree of freedom); calculate a standard error e for an associated slope distribution according to

$$e = 1. / s^c, \qquad (17)$$

where the exponent c was selected for that ensemble; select a slope from the Gaussian number generator; step in the array s positions at that slope, adding to the previously attained height z. The height distribution was further modified. I made the first and last heights equal by removing a straight line. In a fast Fourier transform a quadratic trend is aliased into the low frequencies, so I removed that one of 4 parabolas equally spaced in the array that gave the lowest sum of squared z (usually a very minor modification). I then removed any mean tilt by fitting a straight line to all the heights. Finally, the end of arc was rendered continuous by shifting 1 or a few of the heights. This avoids excess power near the Nyquist frequency. The mean array height was set to zero and the variance adjusted to a selected value. The array was loaded with cos z + i sin z, transformed, and

squared to give Figure 3. This alignment procedure successfully confined the "specular" peak to one frequency bin. The effect of finite σ was established by comparing the power spectrum to that of another array loaded with 0 + i z. The comparison showed essentially identical spectra except for the dc term, provided $\sigma < 0.1$ radian. The ACF is the Fourier transform of the power spectrum, except that the dc term of the array was replaced by the power at the lowest frequency to make the power spectral density function continous there.

For every exponent c the "scattered light" power has a nearly constant region to a tilt g corresponding to 2 to 4 waves in the array length L. The turn-over tilt g governs the decay length for the ACF, which is therefore $L/2\pi$ divided by 2 to 4. At tilts larger than g there follows a power-law regime with slope from -3.4 to -2.0, then a tendency to become a white spectrum. The slope magnitude decreases slightly as the roughness increases and the near-white section becomes shorter. The ACF is almost invariant in shape until $\sigma > 1$; then the decay length falls rapidly. The exponent c = 1 seems natural for a polished surface. A short facet can have a steep slope and a larger facet can have a smaller slope such that the mean slope change per facet is constant. For an exponent c = 0 the slopes all have one distribution and most roughness is contributed by large facets. The "scattered light" at small tilts is enhanced and the descent to large angles is steep, as though big facets are probed at small angles. On the other hand, an extensive white spectrum ensues because unresolvable sharp corners with large slope changes abound. The curve is reminiscent of ground glass. For a large exponent such as c = 10 the big facets are very flat and are separated by short steps that contribute most of the roughness. It is the last case that has a power-law slope closest to -2, and a white spectrum of the slopes, as in a conjecture by Dravins (1978).

I also found it essential that slopes extend down to an unresolvable size. The array can be loaded with the slopes z', transformed, and filtered so that the (nearly) white spectrum of amplitudes is caused to cut off rapidly near half the Nyquist frequency. The back transform gave new slopes z' which were added up as before to give z. Two remarkable consequences follow when all the facets are resolvable. The power spectrum of z shown in Figure 3 straightened from the turnover frequency g to the Nyquist frequency; the new exponent was about half of the steepest slope noted in Figure 3. For c = .01, 1., and 10 the exponent became 1.47, 1.44, and 1.12, respectively. Secondly, the power spectrum became very noisy, especially for increasing exponent c. Beams from facets are strong and average out slowly. Unresolvable scratches spoil those beams in what one would term "a polished surface". Direct measurement with a translating stylus (Elson & Bennett,

1979b) confirms that much surface irregularity on a polished surface involves unresolvable features, smaller than 0.5 μm.

In the above models the ACF had only one characteristic length, e.g., the lag length for decay of the ACF by 1/e. That lag length was from L/10 to L/20, depending on the slope dependence on facet size. It was otherwise a fixed fraction of L. For subintervals no longer than the 1/e lag length, the variance per unit length was constant. For such subintervals the system can be approximated as a stationary random process, just as Keshner (1982) argued. To introduce a second characteristic length such as a facet size I found that the slopes at separations of a facet size must be appreciably correlated. Thus when slopes of a "scratch" were chosen in pairs such that the second slope was almost the negative of the first one, the ACF became the sum of two exponentials. The short decay length was about half of the "scratch" width. Experimental data with such behavior was reported by Elson & Bennett (1979b). Unfortunately, their data analysis and that of Bennett (1976) does not employ the fast Fourier transform and is thus so time consuming as to preclude repeated scans on most samples. The slow decay therefore has fluctuations comparable to the height of the second exponential. Nevertheless it should be noted that their curves pass through zero at a lag of about L/10. A similar criticism applies to Eastman & Baumeister (1974), whose interferometer was described by Eastman (1980); again, their decay length of 0.2 mm was about L/7. Their linear resolution did not detect scratches of finite width; their pixel was 7μm on the test piece.

VI.B. Models of a Grinding Process

The Markov model could be criticized by asking whether any polishing process gives a final height distribution resembling a Markov chain. The Markov chain is efficient in filling an array with numbers, but is it realistic? I therefore simulated a grinding process by a small change in the above procedure. Having picked a facet size s and associated slope, I made that slope positive and began a scratch with two equal facets at a random location in the array. If the groove ran past the end of the array, it was continued at the beginning. The width of the scratch was added up until the net width in grooves was a selected multiple of the array length L. The process started with z = 0. The results at net width = 5L for the same 3 extreme exponents of the facet size-slope relation are shown in Figure 4.

For the exponent c of 1 or less Figure 4 shows nearly the Markov chain result. For a larger exponent c the large and very flat facets become all covered with little scratches; the scattering quickly becomes isotropic. The development with net scratching at c = 1 is shown in Figure 5. The increased scratching causes a small widening of

the isotropic scattering region at high tilts. The changes for the more marginal case of exponent c = 2 are more drastic in Figure 6. A rising variance at small tilts testifies to a gradual growth of low frequency "rumble" in a slow convergence toward the Markov chain result also shown in Figure 6 for comparison. Convergence of this numerical experiment is not asserted nor assured. A real polishing process plausibly scratches differently on high spots or high slopes and involves clumping of the grit. Further experiments could be envisioned.

VII. Two Dimensions in Light Scattering

VII.A. The Isotropic Aperture

A good polishing process leaves tiny scratches, just as many in every azimuth. Disappointment is known even in coronagraphic grade optics (Smartt, 1979). Then we should expect an exponential-like ACF would be measured in any direction of choice. The Fourier transform of an azimuthally symmetric function was treated by Born & Wolf (1970, pg. 395). An azimuthal integration yields a zero order Bessel function. The remaining radial integration is then approximately the Hankel transform of an exponential (a small error is made by extending the upper limit of integration from 2D to infinity). The power spectral density function becomes (Stover, 1975; Church et al., 1977)

$$W_2(r) = 2\pi a^2 g^{-2}(1 + (r/g)^2)^{-3/2} \qquad (18)$$

where a is an rms height such that at normal incidence we have

$$\sigma = 2\,ka \qquad (19)$$

and r is a spatial frequency,

$$r^2 = p^2 + q^2 \qquad (20)$$

The appearance of g^{-2} in (18) is needed from a units point of view; it will cancel two spatial frequencies when a double integration over spatial frequencies gives the total variance a^2. For a scattering angle large enough so that r exceeds g (a tilt of 2 or more waves in the aperture D) the power falls off in angle as θ^{-3}.

VII.B. The Kormendy Small Angle Region

The symmetry giving (18) is plausible on axis where p = q = 0. Off axis in any direction the symmetry is broken by the effective grating whose period d is given by (1). For small angles and small N = 2 D/d we must ask whether the averaging of the aberrations z(x,y) over the N strips of the grating is in some sense statistically independent of the orientation of the graing. If so, stating that symmetry is roughly equivalent to giving N equations of condition. At wide angles we find the fall-off of the power spectrum of z goes as θ^{-2}, which is N^{-2}, and that it was associated with a one-dimensional ACF, i.e.,

dominance of linear scratches in fine details of $z(x,y)$. The phase contrast photographs of Smartt (1979) reveal a maze of linear features extending across most of the aperture. His phase-shifting dot size was too large to visualize aberration features not having one dimension smaller than about 2 mm in a scene of 3-cm size. His reference dot was 120 μrad in radius, which might miss a θ^{-3} region like that of Kormendy. Since at smaller angles the bigger features of $z(x,y)$ are not apt to be linear, it is reasonable to expect azimuthal symmetry. In turn, N equations of condition then change the sorting of variance from N^{-2} to N^{-3} in angle, i.e., θ^{-3}.

Test data on large optics abound, but I do not know that any have been analyzed to give an ACF. In V. we mentioned the exponential-like ACF of Rezette for sanded glass with a method that uses all of the small-angle scattered light. The work of this section shows that a θ^{-3} region, if it existed, would not spoil an exponential ACF.

VIII. Wavelength Dependence of the Scattering

VIII.A. Differential Light Scattering

We noted in (15) that wide-angle scattered light intensity is proportional to k. The powers of k contained in the spatial frequency p (see (14)) and the radian measure of roughness σ (see (19)) exactly cancel in the ratio $(\sigma/p)^2$. If the angular exponent were $2-\epsilon$, we would expect an overall wavelength dependence proportional to $k^{1+\epsilon}$ at wide angles.

To generalize to other cases simply, we use the ratio of powers into the zero and first order for a sinusoidal grating of amplitude a_j (Stover, 1975) using s-polarized light,

$$\text{s-ratio} = (ka_j)^2 \cos\theta \cos\psi . \qquad (21)$$

The factor a_j^2 is one term in a variance sum whose total is a^2. We would like to deal with a power spectral density function in the case of very many gratings. To that end we multiply the numerator and denominator of (21) by the differential of (14),

$$\Delta p = k \cos\theta \, \Delta\theta \qquad (22)$$

so that for one dimension the increment of light power is

$$P_1 = P_o \, f \, k^3 \, W_1(p) \, \Delta\theta, \qquad (22)$$

where

$$W_1(p) = 2\pi^{-1} a^2 g/(g^2 + p^2) \qquad (24)$$

and f was given by (16). To go on to 2 dimensions we will need another spatial frequency increment Δq in the denominator, another power of k in the numerator; the interval $\Delta\theta$ becomes the solid angle Ω. More exact treatments of the transition from $W_1(p)$ to $W_2(p,q)$ were given by Elson & Bennett (1979b) and Wang & Wolfe (1983). Also, Church (1979) notes that they are related by an integral transform. When these

thoughts are applied to (18), we note that the Kormendy θ^{-3} scattering region also has a wavelength dependence as k because in the ratio k^4/r^3 there are 3 powers of k that cancel.

VIII.B. Total Integrated Scatter

The sum over all angles gives the total integrated scatter (TIS), which is approximately σ^2 for polished surfaces. The wavelength dependence of TIS is then k^2 according to (19) and was confirmed in a variety of experiments (Elson & Bennett, 1979a for a review) from the ultraviolet to the near infrared. Summing over a finite range of angles is the experimental reality and was discussed by Church et al. (1979). They emphasized the importance of θ^{-2} scattering to the k^2 finding for TIS. For the best photometry we note that the focal plane aperture size must not be changed. The escaping energy is not white.

VIII.C. Other Effects and Problems

A key experiment because of the wavelength range from 0.633 to 10.6 μm was that of Wang & Wolfe (1983), from which Figures 7 and 8 are taken. Figure 7 gives the scattered power per unit solid angle divided by the input power (BRDF) as a function of the logarithm of sin θ - sin Ψ. In Figure 8 similar data are plotted as the function $W_1(p)$ in (24), the angular factors having been divided out. The curves are shifted vertically to remove the k-dependence of (15) and shifted left to right as required for p as in (14). The near coincidence of the curves supports the theories of rough surface scattering discussed so far.

A second look at Figure 8 reveals several problems. The data for 0.63 and 3.39 μm have a slope near -2 (closer to -2.2) over 3 decades in p, but the data for 10.6 μm in an overlapping p-range have a slope of only -0.7. At grazing emergence for every wavelength the power exceeds that predicted by theory. This "curl up" feature was also found by Stover (1975) and Thomas (1980). In that same angular region Stover et al. (1983) and Thomas also found that the scattered power had a dependence on polarization state not quite in agreement with theory. Stover and Hourmand (1984) carefully sought experimental errors that might cause "curl up", but defended the reality of the phenomenon.

The deviations just noted suggest micron-sized particles on the mirror surfaces. Unlike the usual telescope condition, the samples studied by the above authors were very clean and smooth. The sample of Figure 7 had a directly measured roughness of 1 nm. At 10.6 μm this gives $\sigma = 10^{-3}$ and total scattering per sr is $10^{-6}/0.03$ sr, the observed BRDF. The shorter wavelengths have both a larger σ (in

radians) and a narrower region of scattering, which raises their BRDF. I estimate that one particle per 10 mm^2 having radius b = 1.7 μm would give the observed level of scattered light at 10.6 μm. It would take a powerful search routine to find such a low density of small particles, but it is available commercially in the form of a laser beam scanner used by the solid-state device manufacturers. Those scanners integrate scattered light in a large solid angle. If "curl up" is caused by particles, then it follows that a scanner based on the ratio of light launched at grazing emergence to that launched closer to normal emergence will discriminate against scratches. For astronomy a monitor of particles on the primary mirror need not scan. It would be enough to provide a laser beam a radian or more off axis and detectors that view the illuminated region from two angles.

Against the suggestion of contaminating particles might be the apparent wavelength dependence of scattering from the particles. The 3.4 μm data were not swamped by the radiation that changed the 10.6 μm data drastically. Therefore the excess radiation must be nearly independent of k. Responding to one power of k made the 3.4 μm light from surface roughness substantially stronger than its competitor. The net scattering and absorption by a Junge distribution of particle sizes

$$N(b) \propto b^{-m} \qquad (25)$$

is often well-represented for visible light (Trakhovsky et al., 1982) by the Angstrom law

$$\gamma = C\, k^{m-3} \qquad (26)$$

which is independent of k for an urban type of aerosol. Nevertheless, this retort has full effect in the absence of direct data, so I shall refer to the "particles" as "singular points".

The interest in singular points of a few μm or less is that the condition kb = 1 is also the condition (van de Hulst, 1957) for the onset of efficient scattering for an isolated particle. It also marks the boundary between Rayleigh scattering into a wide solid angle and the forward scattering of larger particles. The excess radiation would be emphasized at grazing emergence where the scattering by surface roughness must vanish. In addition, the Rayleigh regime imparts new polarization effects.

A physical theory of particles on a surface would not be easy. In the rough surface theories the radius of curvature must exceed the wavelength (Beckmann & Spizzichino, 1963, pg. 20; Elson & Bennett, 1979a). When this condition is violated, one expects a change from diffraction that is predominantly forward to a wider region. My computer models depend on this forward scattering condition. Secondly, particles on a mirror are not isolated; they should be treated as dipoles. If they have resisted removal procedures, they may be closer to hemispheres than spheres. If they are under a metallic coating,

that coating is probably inhomogeneous.

Direct evidence for singular points on coronagraphic mirrors is the field of star-like points distributed evenly across the surface reported by Smartt (1979) with dark-field illumination. He judged that a fainter but uniform distribution of light dominated the net scattering, but quantitative work is needed. He found some singular points modulated by changes in geometry, so he implicated crystalline-like facets.

A practical concern for telescopes is that surface scattering by particles, even if it strongly mimics mirror roughness scattering, can substantially modulate the encircled energy for a 100 or 200-ring focal plane aperture. A change in σ^2 on the level of parts per thousand and on the time scale of hours is very serious. Yet it is possible, since atmospheric particles acquire liquid water mantles at high humidity. The development of a particle monitor based on the "curl up" discussed above might permit the assessment of encircled energy modulation.

I am indebted to A. M. Dainty for the reminder of the theorem of Bracewell. Part of this research was supported by the Computer Center of the University of Arizona.

IX. References

P. Beckmann & A. Spizzichino (1963), The Scattering of Electromagnetic Waves from Rough Surfaces, Pergamon Press, Oxford.
P. Beckmann (1973), IEEE Trans. Antennas Prop. AP21:169.
J. M. Bennett (1976), Appl. Optics 15:2705.
M. Born & E. Wolf (1970), Principles of Optics, 4th ed., Pergamon Press, Oxford.
R. Bracewell (1978). The Fourier Transform and Its Applications, 2nd ed., McGraw-Hill, N. Y., p. 143.
E. L. Church, H. A. Jenkinson, & J. M. Zavada (1977), Optical Engrg. 16:360.
E. L. Church, H. A. Jenkinson, & J. M. Zavada (1979), Optical Engrg. 18:125.
E. L. Church (1983), S. P. I. E. 429:86.
A. M. Dainty (1984), J. Geophys. Res. 89:3172.
J. L. Doob (1953), Stochastic Processes, Wiley, New York.
D. Dravins (1978), Appl. Optics 17:404.
J. M. Eastman & P. Baumeister (1974), Opt. Commun. 12:418.
J. M. Eastman (1980), Optical Engrg. 19:810.
J. M. Elson & J. M. Bennett (1979a), Optical Engrg. 18:116.
J. M. Elson & J. M. Bennett (1979b), J. Opt. Soc. Am. 69:31.
T. Gehrels, D. L. Coffeen, & D. Owings (1964), Astron. J. 69:826.
J. W. Goodman (1975), Laser Speckle and Related Phenomena, Springer-Verlag, Berlin.

J. E. Harvey (1976), Ph. D. Thesis, Univ. Ariz., Tucson, AZ.
J. E. Harvey (1977), S. P. I. E. 107:41.
C. E. KenKnight (1977), Icarus 30:422.
M. S. Keshner (1982), Proc. IEEE 70:212.
I. R. King (1971), P. Astron. Soc. Pacific 83:199.
J. Kormendy (1973), Astron. J. 78:255.
A. Papoulis (1962), The Fourier Integral and Its Applications, McGraw-Hill, New York.
J. Piccirillo (1973), P. Astron. Soc. Pacific 85:278.
A. K. Pierce (1954), Astrophys. J. 120:221.
J. O. Porteus (1963), J. Opt. Soc. Am. 53:1394.
R. O. Redman & E. G. Shirley (1938), Mon. Not. Roy. Astron. Soc. 98:613.
Y. Rezette (1979), S. P. I. E. 201:50.
R. S. Sayles & T. R. Thomas (1978), Nature 271:431.
R. N. Smartt (1979), S. P. I. E. 190:58.
J. C. Stover (1975), Appl. Optics 14:1796.
J. C. Stover, S. A. Serati, & C. H. Gillespie (1983), S. P. I. E. 429:96.
J. C. Stover & B. Hourmand (1984), S. P. I. E. 511, to be published.
D. A. Thomas (1980), Ph. D. Thesis, Univ. Ariz., Tucson, AZ.
E. Trakhovsky, S. G. Lipson, & A. D. Devir (1982), Appl. Optics 21:3005.
H. C. van de Hulst (1957), Light Scattering by Small Particles, Wiley, New York.
Y. Wang & W. L. Wolfe (1983), J. Opt. Soc. Am. 73:1596.
Y. Wang & W. L. Wolfe (1984), J. Opt. Soc. Am. A1:783.
A. T. Young (1974), in Methods of Experimental Physics, vol. 12B, Astronomical Photometry, Academic Press, New York, pg. 95.
H. Zanstra (1943), Mon. Not. Roy. Astron. Soc. 103:265.

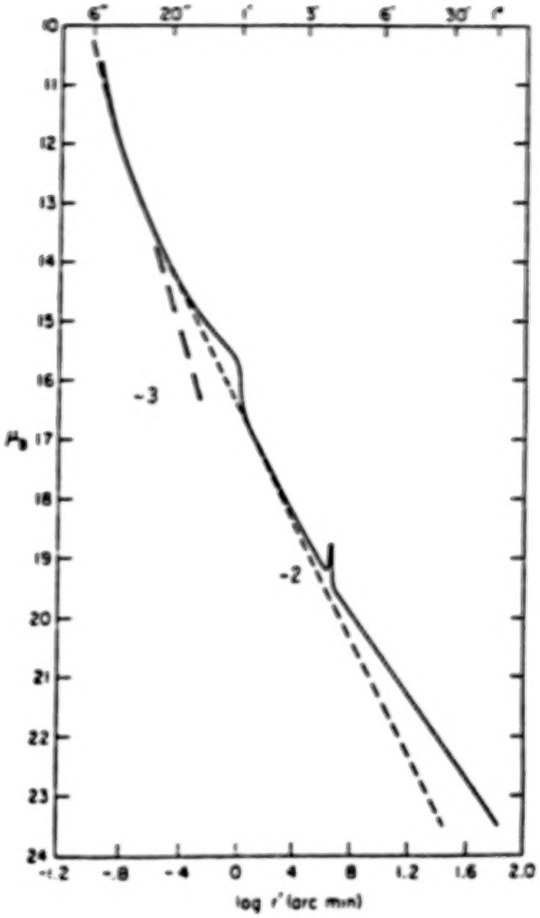


Figure 1. The data from Kormendy (1973) in stellar magnitudes per square arc-second show power law decreases from inverse cube to just slower than inverse square laws.

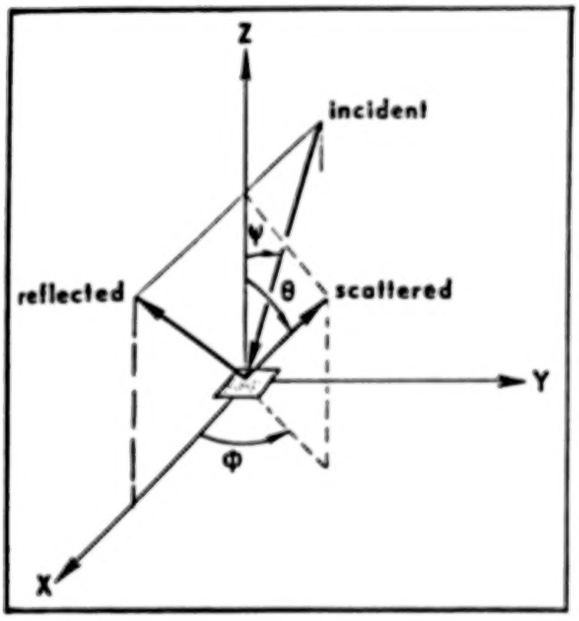


Figure 2. The scattering geometry discussed in the text has light incident in the zx plane and light emerging so that spherical polar coordinates θ and ϕ are used.

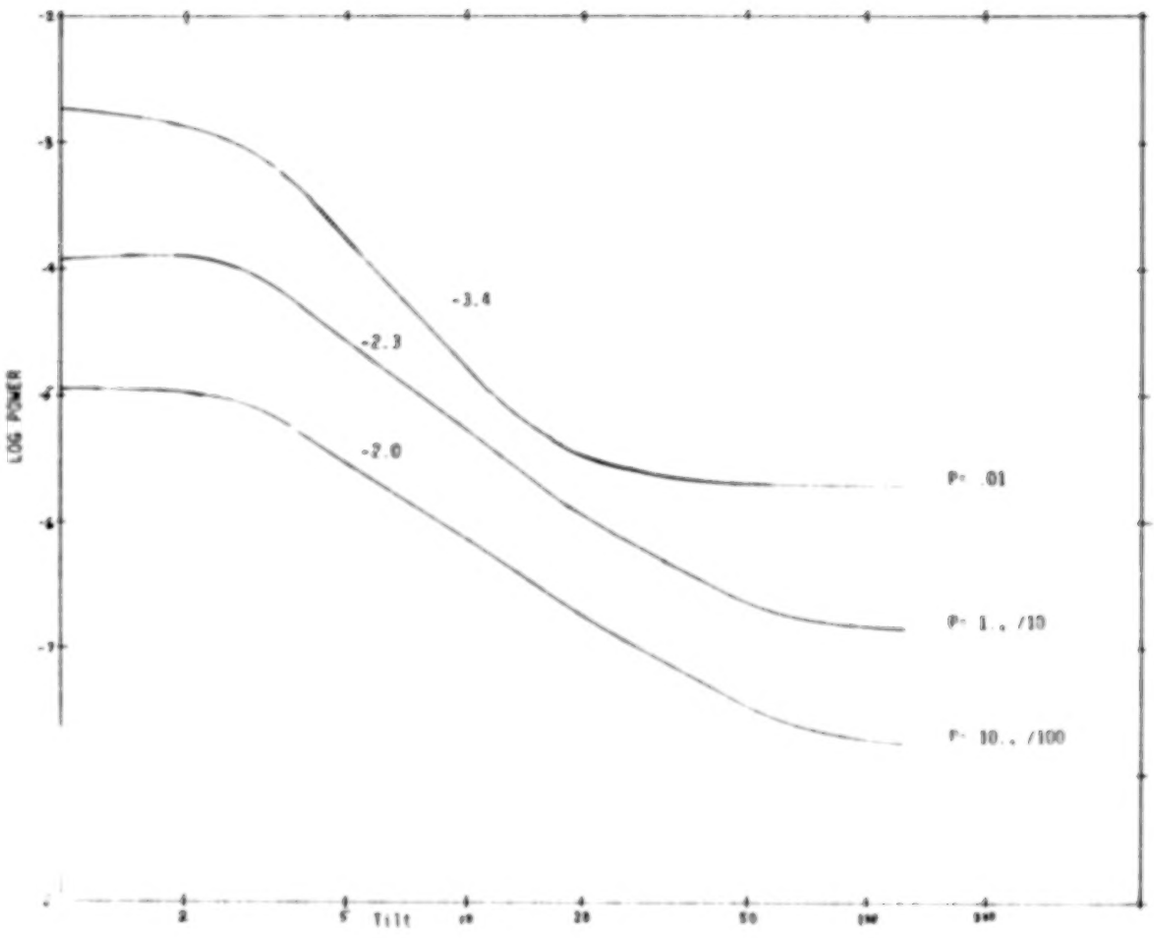


Figure 3. A Markov model with facets whose slopes are inversely related to size with power p gives the aberration power spectrum shown for various spatial frequencies, i.e., tilt of the scattered wavefront.

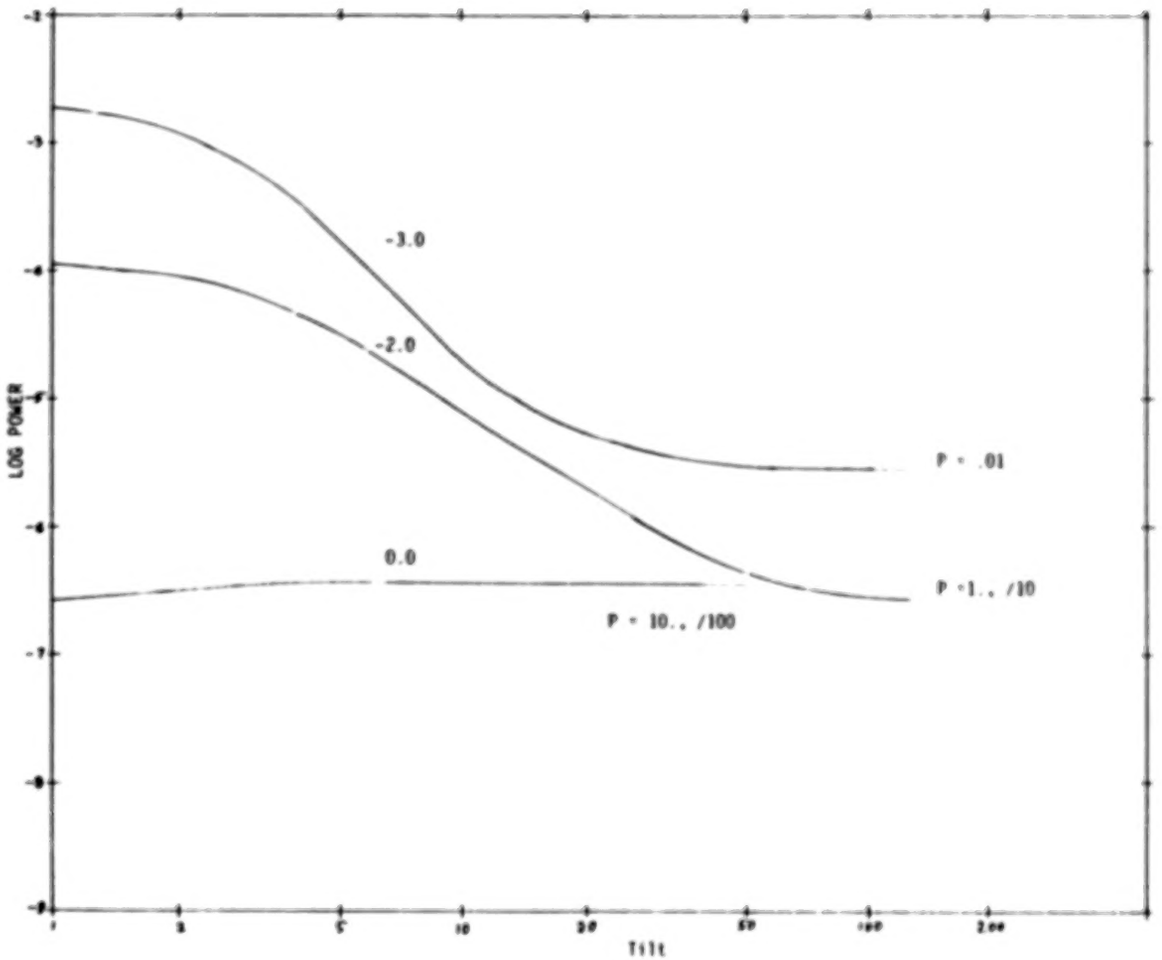


Figure 4. A model of polishing with scratches having facets whose slopes are inversely related to the scratch width is shown at the stage that net scratch width is 5 times the sample width.

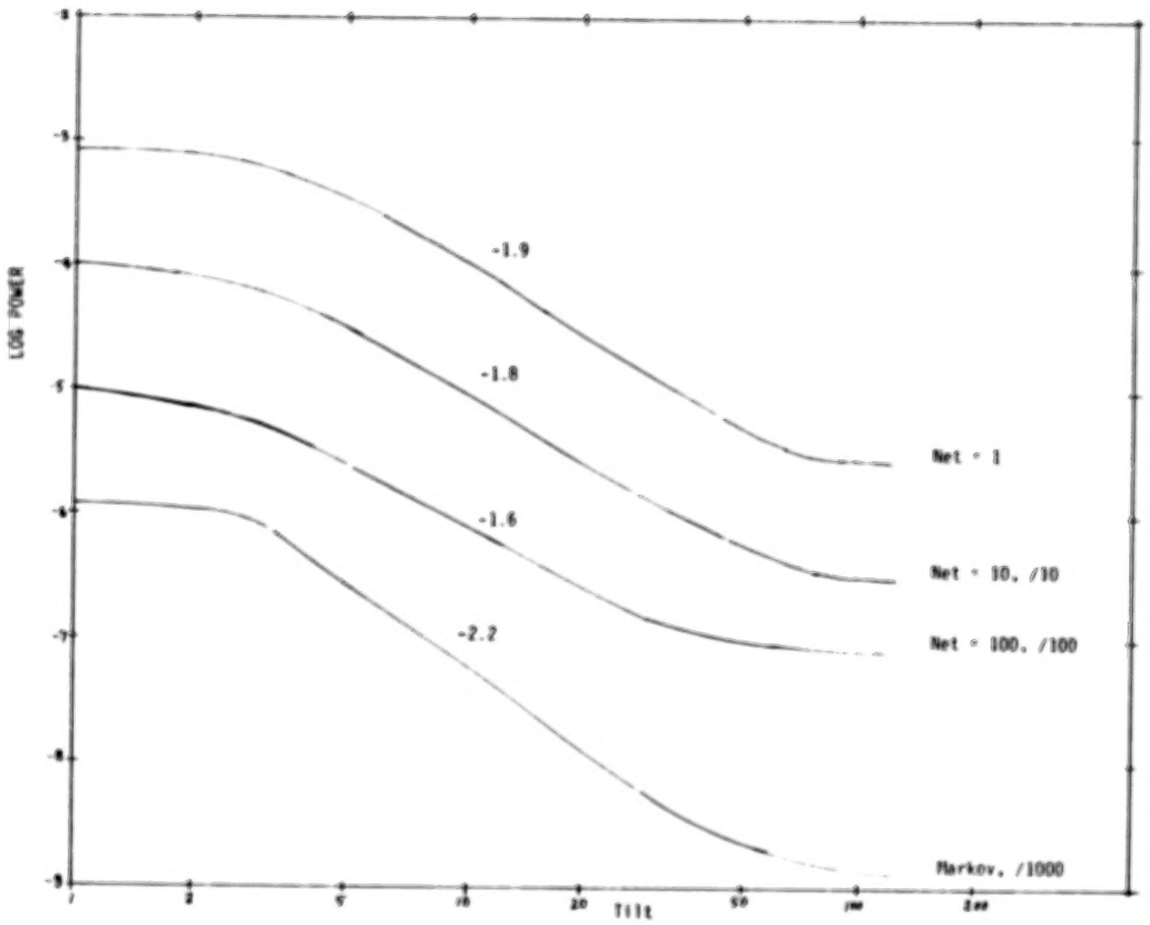


Figure 5. The model of polishing with slope- size exponent p = 1 gives these power spectra at the net scratching indicated. The Markov result is shown for reference.

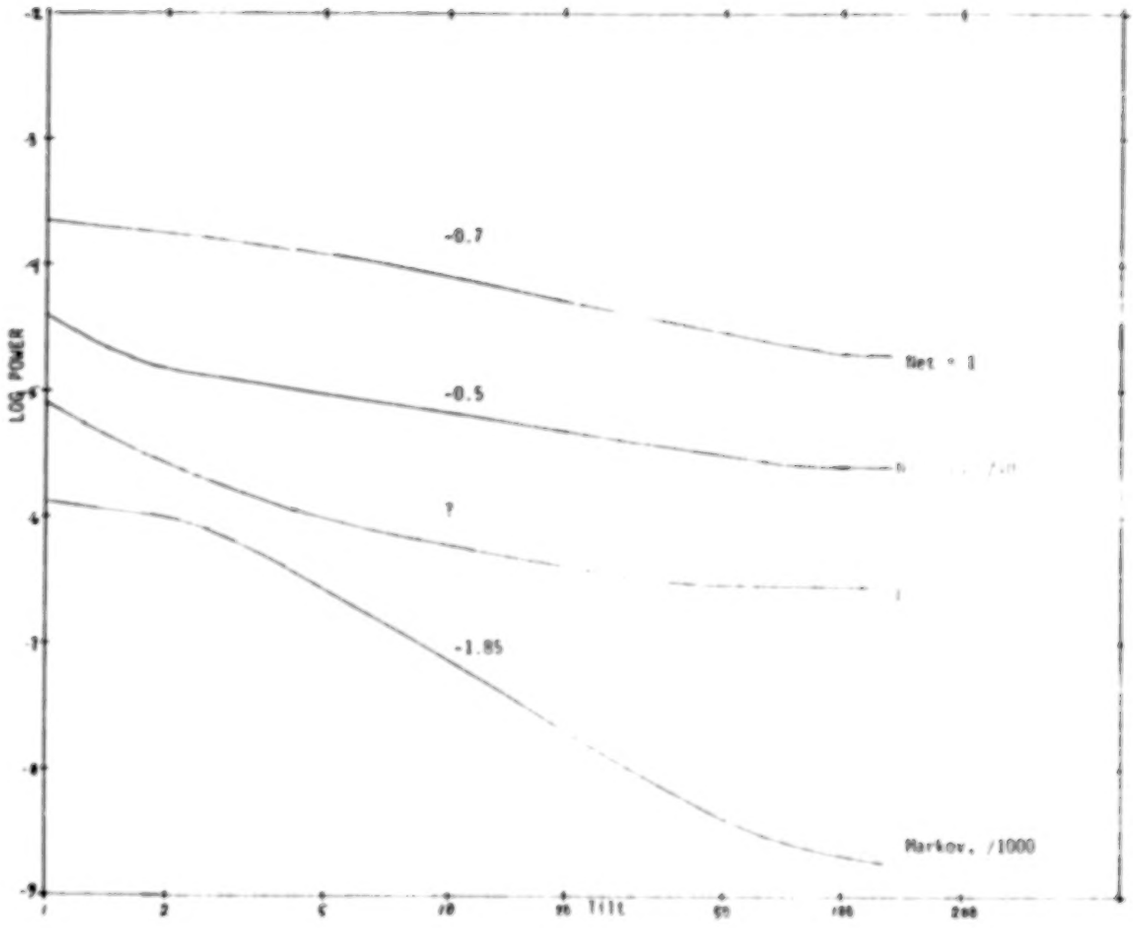


Figure 6. Caption as in Figure 5 except that the power-law exponent p = 2 corresponds to big facets at very flat slopes, which never accumulate little scratches.

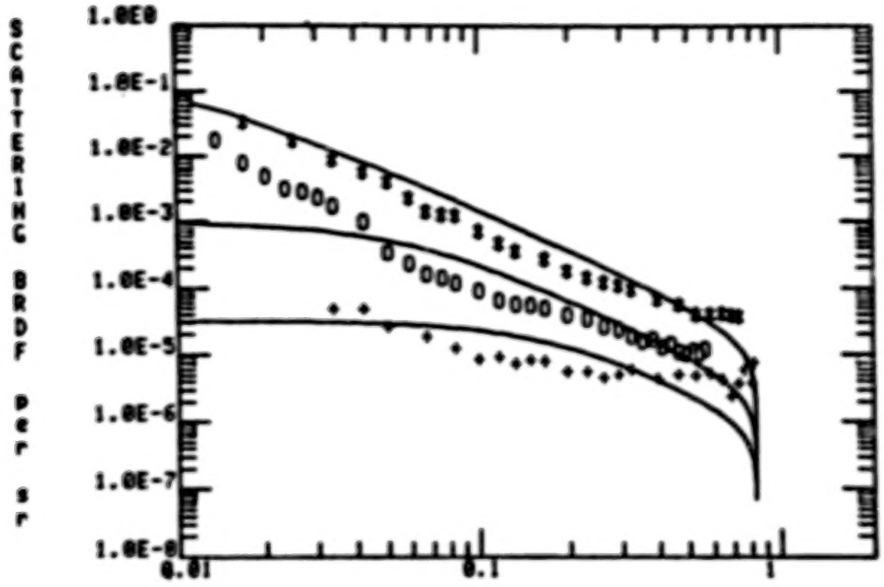


Figure 7. Relative scattering power for a very smooth sample at 0.633 um (*), 3.39 um (0), and 10.6 um (+) for 10 degrees incidence and scattering parameter sin θ - sin ψ from 0.01 to grazing emergence, from Wang & Wolfe (1983).

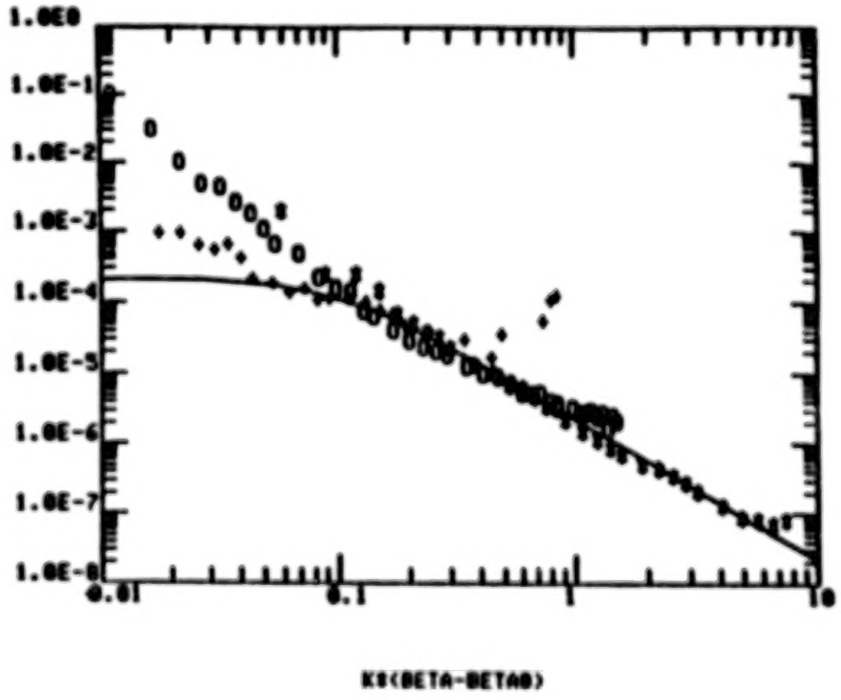


Figure 8. One dimensional power spectra for wavelengths as in Figure 7, -30 degree incidence, against spatial frequency in inverse microns, from Wang & Wolfe (1983).

D20 ; N85-17912

APPLICATION OF FIBER OPTICS AND COMPOUND COLLECTORS TO PHOTOMETRY

Stephen D. Fantone
Optikos Corporation
85 Holland Street
Saugus, MA 01906

Conventional photometric equipment typically contain an assortment of optical components: lenses, mirrors, and filters. Over the past fifteen years a number of new optical components have become available but have not found widespread use as components in stellar photometers. This paper considers two such components: fiber optics and compound collectors, and provides suggestions for their use in photometric hardware. This is followed by a section on system issues and an illustrative example is provided.

Fiber Optics

The basic concept underlying fiber optics is the phenomenon of total internal reflection. This lossless reflective process allows for the propagation of a confined beam down a dielectric structure with total loss determined by fundamental material limitations. Figure 1 depicts the geometry of an optical fiber.

For photometric applications we are interested in transfering flux from one point to another. Therefore, it is of no advantage to use single-mode fibers of the type being used in telecommunication systems. Multi-mode fibers are commonplace and are readily available in two sizes, 50 and 125 micron core diameters.

The numerical aperture of an optical fiber is determined by the refractive indices of the fiber core (N1) and the fiber cladding (N2). That is,

N.A.=numerical aperture=sqrt($N1^2 - N2^2$) .

For example if

N1=1.50,
N2=1.46,
α =13.36 degrees, and
θ =20.28 degrees, then
NA=.346 corresponding to an F/#=1.44 .

Ideally there is no reflective loss at the core-cladding interface and were it not for absorption, scattering losses or mode conversions the fiber optic structure would be a perfect

flux transfer device. There are four major sources of flux loss in a fiber:

1. Absorption by the dielectric material,
2. Rayleigh scattering due to density fluctuations and hence refractive index variations,
3. Scattering due to rough surface boundaries between core and cladding, and
4. Bending losses due to mode conversion losses within the fiber.

Most fiber optics are made from silicate based glasses. Figure 2 depicts the losses per kilometer of length in a typcal low loss optical fiber. Except for the residual absorption due to water ("OH" peaks), the losses in a straight fiber are almost entirely due to Rayleigh scattering.

Within the fiber the intrinsic absorption is a very stable characteristic, thus even though the absorption is dispersive it is not likely to vary over reasonable periods of time (decades) when use in a reasonable environment (temperatures considerably below the glass transition point and low exposure to radioactivity).

Bending losses are more difficult to control. There are two types of bending losses. The first is the most obvious; if a fiber is bent with a radius of curvature of several centimeters, some of the light will "leak" from the fiber. This is readily observable in common optical fibers. Calculation of this bending loss is very complicated but a rule of thumb is that if the bend radius is greater than 1000. times the core diameter the loss is insignificant for fiber lengths up to 10,000. core diameters. It is therefore important to avoid kinks and tigth bends in the fiber. Figure 3 depicts the transmission p(z)/p(0) of a fiber for a family of bend radii (R) to core diameters (p) ratios for various fiber length (z) to core diameter (p) ratios. The second bending loss in fibers is termed micro-bending losses. These are due to small periodic modulations in the core-cladding interface. These modulations are introduced during fiber manufacture and cabling and can be avoided using proper techniques.

The important feature of optical fibers is that they allows separation of the signal detection and amplification equipment from the telescope and are intrinsically stable and repeatable over long periods of time.

Compound Collectors

Compund collectors are members of a family of flux collection devices termed non-imaging concentrators. Most common flux collection devices consist of a mirror or lens which images a source onto a detector. To contrast this, consider the function of a a conical reflector placed in front of a detector. No image is formed. The cone does limit the field of view of the detector and does redirect flux into the detector that would normally

miss. There exists a class of these non-imaging detectors that in 2 dimensions are "perfect" collectors. By perfect, we mean that all of the flux incident on the entrance aperture within a limited field of view eventually (after many reflections) are directed into the exit aperture of the collector and onto a detector. This collector is depicted in Figure 4a and its imagery depicted in Figure 4b. The collector in 2 dimensions consists of 2 parabolic curves with axes symmetrically displaced from the optical axis of the collector. When generalized to 3 dimensions the collector is not quite perfect, but the collection efficiency and rejection of out -of-field souces is exceptional. Figure 5 depicts the collection efficiency for several 3 dimensional compound parabolic collectors.

There are number of variations on this same theme. The collectors can be made of a refractive material. This decreases the overall size of the collector and allows the use of total internal reflection to achieve high reflectivity in the colllector. The collectors can be ganged such that the output for one is the input for another. It is important to keep in mind, however, that the ratio of input aperture diameter to output aperture diameter must equal the ratio of maximum output ray angle to maximum input ray angle (field coverage). This is a consequence of the Lagrange theorem and the second law of thermodynamcs.

Compound collectors can serve a useful flux collection and averaging devices in photometers. Example - since a compound collector has uniform flux transfer properties across its entrance aperture, the throughput of the collector is not affected by motion of an image in the aperture. That is, if a compound collector is placed in a stellar image plane, the output of the collector is independent of image motion so long as the image remains within the entrance aperture.

System Considerations

In utilizing both fibers and compound collectors in a photometric system it is important to keep in mind the flux path through all of the optics. Clearly, if we are using a compound collector as an input device to an optical fiber it is important that the numerical aperture of the fiber exceed the numerical aperture of the collector. This will allow for assembly tolerances when mating the two components. Similarly, it is important that the field-of-view of the compound collector allow 100% collection of the flux from the focussing element before it. That is, the numerical aperture of the primary collecting optics must be less than the numerical aperture of the collector.

A focal plane can be easily populated by an assembly of compound collectors and fibers. Each fiber would then lead to its own signal detection and processing electronics. Figue 6 depicts a simple fiber optic photometer. Such a system has a number of additional advantages. All heavy, bulky, and heat producing equipment can be located off the telescope in a more appropriate

area. The focal plane can be densely packed with compound collectors sized to a small multiple of the seeing disk size. Multi-channel analysis allows for more simultaneous measurements and hence more directly comparable data. The focal plane can easily be reconfigured to look at a different stellar field.

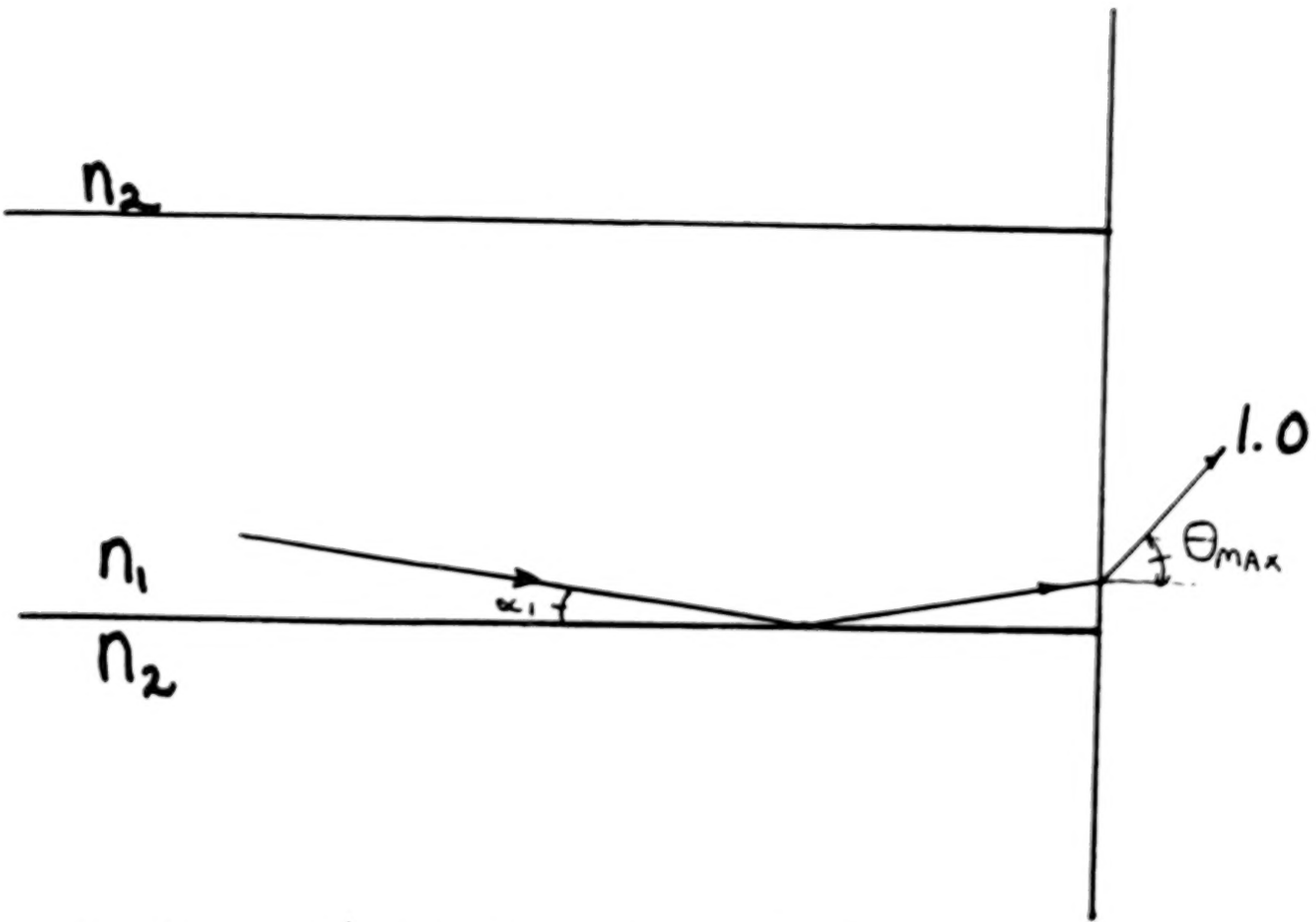
Conclusion

The development of new optical components will allow the construction of a new class of stellar photometers with previously unobtainable flexibility and reliability. Consideration of both compound collectors and fiber optic flux transfer should lead to more precise and repeatable photometric data.

References

Winston, R. and Welford, W.T., The Optics of Non-imaging Concentrators, Academic Press, NY, 1978

Marcuse, D., Light Transmission Optics, Van Nostrand Reinhold Co., NY, 1972



$$N.A. = N \sin \alpha_{MAX} = n_1 \sin \alpha_1$$

$$= 1.0 \sin \theta_{MAX}$$

FIGURE 1

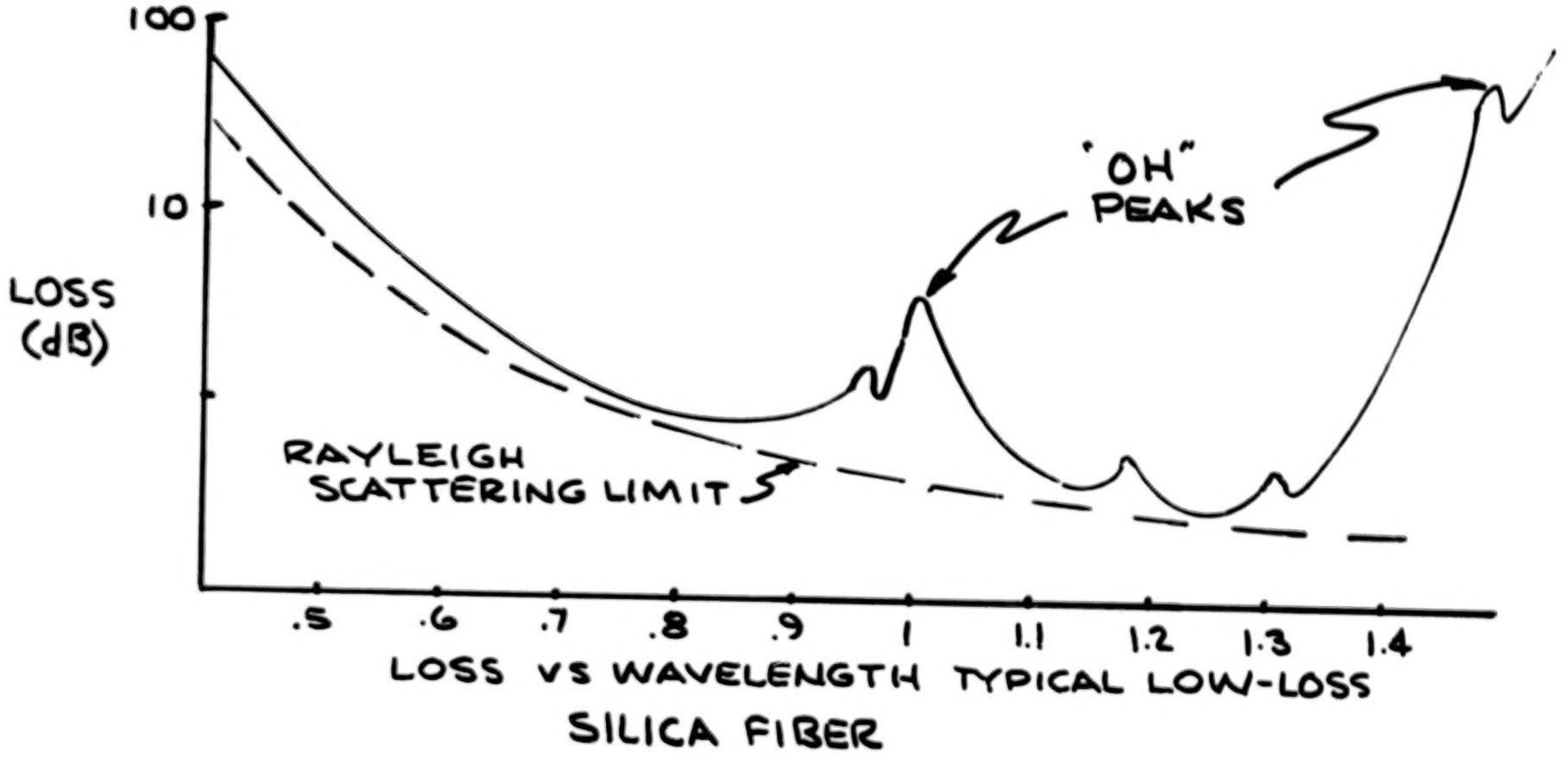

100
10
LOSS
(dB)
"OH"
PEAKS
RAYLEIGH
SCATTERING LIMIT
.5
.6
.7
.8
.9
1
1.1
1.2
1.3
1.4
LOSS VS WAVELENGTH TYPICAL LOW-LOSS
SILICA FIBER

FIGURE 2

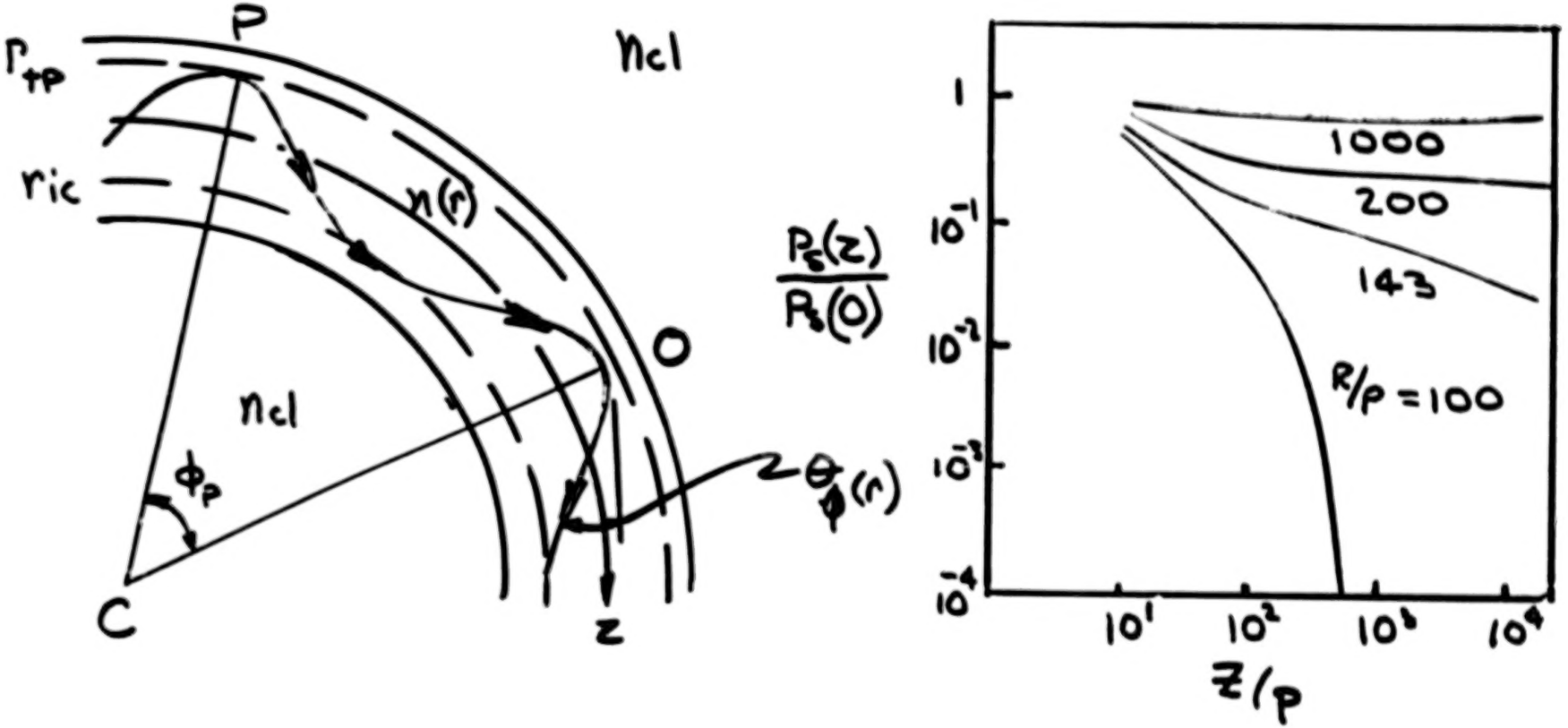

P
r_{tp}
r_{ic}
n_{cl}
$n(r)$
n_{cl}
ϕ_p
C
O
z
$\theta_\phi(r)$
$\frac{P_s(z)}{P_s(0)}$
1
10^{-1}
10^{-2}
10^{-3}
10^{-4}
1000
200
143
R/p = 100
10^1
10^2
10^3
10^4
z/p

FIGURE 3

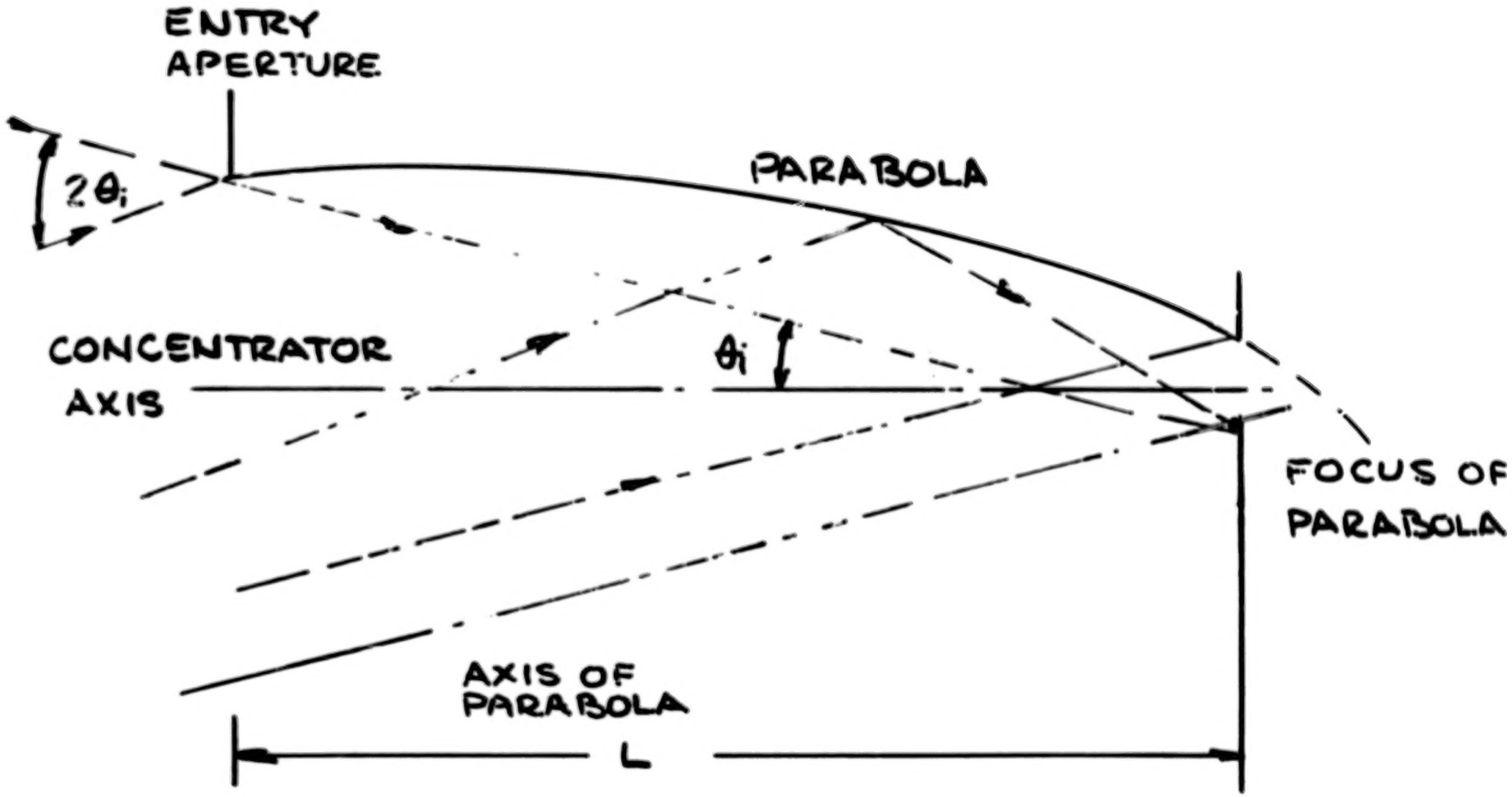

ENTRY
APERTURE
$2\theta_i$
PARABOLA
CONCENTRATOR
AXIS
θ_i
FOCUS OF
PARABOLA
AXIS OF
PARABOLA
L

FIGURE 4A

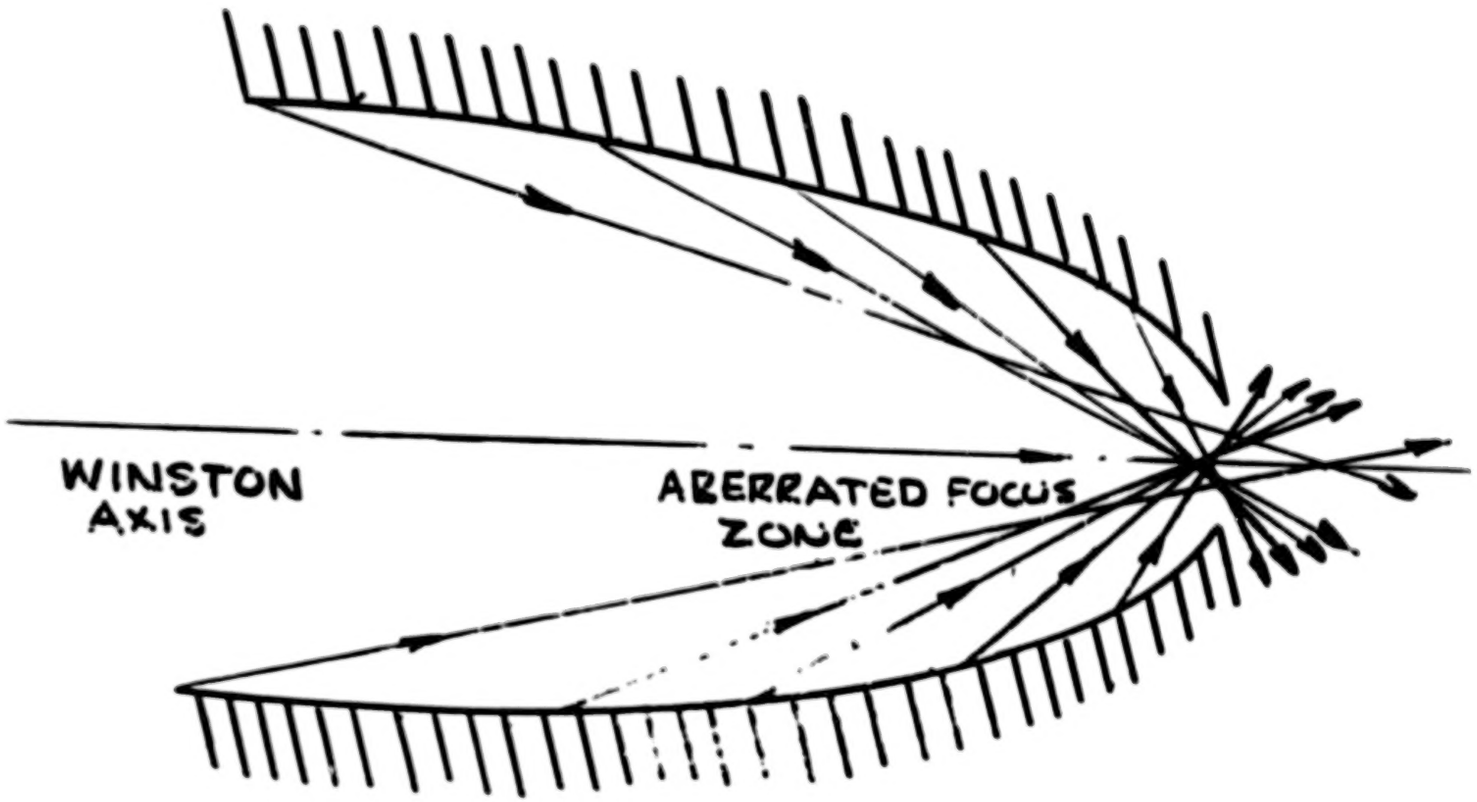


FIGURE 4B

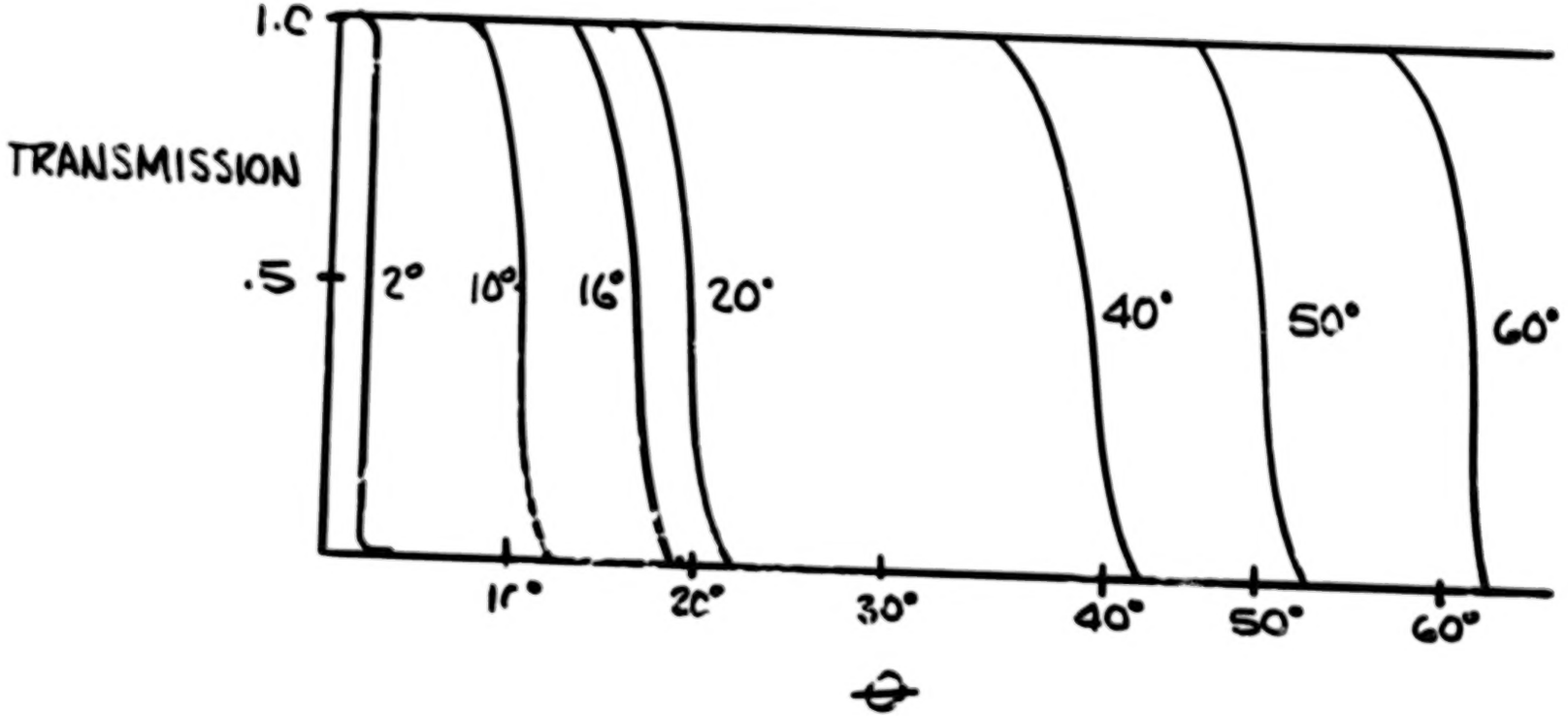


FIGURE 5

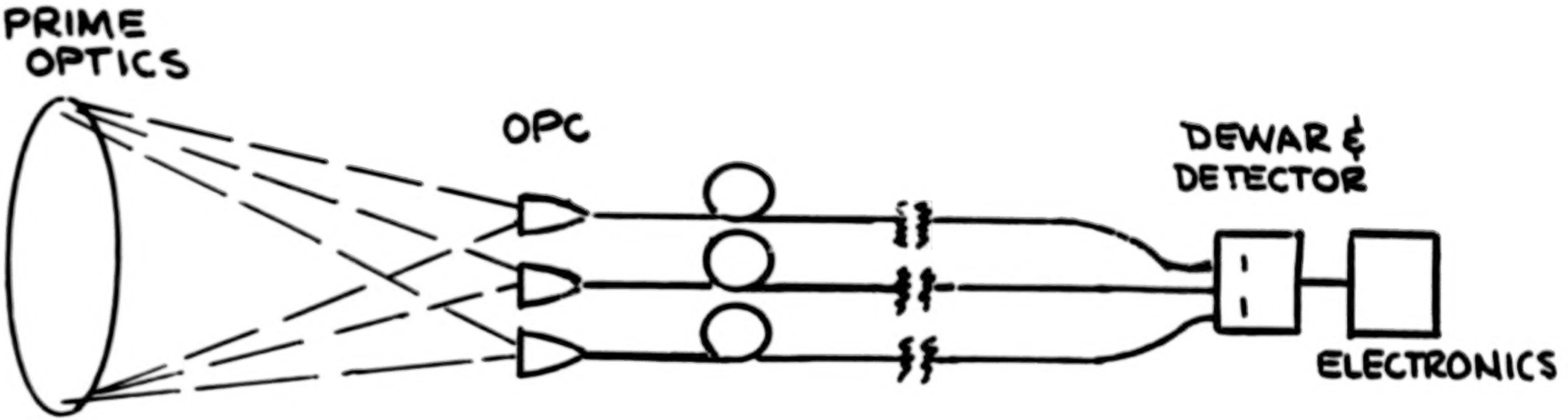


FIGURE 6

ASTRONOMERS WORKSHOP PARTICIPANTS

Donald J. Angione
San Diego State University
Astronomy Department
San Diego, CA 92182
619-265-6183

James Baker
Center for Astrophysics
60 Garden St.
Cambridge, MA 02138
617-495-1693

Frank L. Beale
San Diego State University
Astronomy Department
San Diego, CA 92182
619-265-6674

William Borucki
NASA Ames Research Center
Bldg. 245-3
Moffett Field, CA 94035
415-694-6492

Anita Cochran
University of Texas at Austin
Astronomy Department
RLM 15.308
Austin, TX 78712
512-471-4461

William Cochran
University of Texas at Austin
Astronomy Department
RLM 15.308
Austin, TX 78712
512-471-4461

S. Djorgovski
University of California, Berkeley
Astronomy Department
Berkeley, CA 94720
415-642-3958

Stephen D. Fantone
Optikos Corporation
85 Holland Street
Saugus, MA 01906
617-231-2044

Eric Fossat
Observatoire de Nice
B P 139
F-06003 Nice Cedex
France
93-519100 X 1345

Jon Geist
National Bureau of Standards
Room A 314 MET
Washington, DC 20234
301-921-2191

Hugh Hudson
University of California, San Diego
UCSD C-011
La Jolla, CA 92093
619-452-4476

Bernard V. Jackson
University of California, San Diego
Dept of EE/CS C-014
La Jolla, CA 92093
619-452-3358 / 2022

Charles KenKnight
7740 E. Nicaragua Drive
Tucson, AZ 85730
612-790-1712

Donald W. Kurtz
University of Cape Town
Department of Astronomy
Rondebosch 7700, South Africa
69-8531 X 629

Wes Lockwood
Lowell Observatory
Box 1269
Flagstaff, AZ 86002
602-774-3358

Robert S. McMillan
University of Arizona
Lunar and Planetary Lab
Space Science Building
Tucson, AZ 85721
612-621-6968

Warren Rosen
Vassar College
Poughkeepsie, NY 12601
914-452-7000

Fredy Rufener
Geneva Observatory
CH 1290 Sauverny
Switzerland
22-55-26-11

Russ Schaefer
National Bureau of Standards
B-306-MET
Washington, DC 20234
301-921-2191

Andrew T. Young
San Diego State University
Astronomy Department
San Diego, CA 92182
619-265-5817

END

DATE

FILMED

APR 2 1985